Journal of Experimental Psychology

Journal of
Experimental Psychology

VOL. II, No. 5. OCTOBER, 1917

A REPETITION OF EBERT AND MEUMANN'S PRACTICE EXPERIMENT ON MEMORY[1]

BY H. B. REED

University of Idaho

Reviewers of Ebert and Meumann's practice experiment upon memory are generally agreed that the experiment must be repeated before their conclusion can be accepted as valid for the facts of the experiment. Lack of space prevents me from taking notice of the reviews and criticism that have been made on their work. The references of Müller, Sleight, and Thorndike cited below will be found valuable in this connection.

Suffice it to say that the general conclusion drawn from the results was that a special training of the memory improves it in general. But since the amount of transfer from the special training to the tests was not equal in all the tests, a particular conclusion was drawn in that the amount of transfer follows the law "that the special memories are improved exactly to the degree that the learning materials and learning methods are related to the specially trained memory" (p. 200). These conclusions have not been accepted and have been criticized among others by G. E. Müller[2] and R. Wessely[3] in Germany, by W. G. Sleight[4] in England, and W. F. Dearborn[5] and E. L. Thorndike[6] in America.

So far as is known to the writer, Meumann took no notice of his critics except G. E. Müller whom he ridicules for his

[1] *Archiv. f. ges. Psychol.*, 4, 1904, 1–232.
[2] *Zeit. f. Psychol.*, 39, 1905, 111–125.
[3] *Neues Jahrb. f. Päd.*, 16, 1905, 296–309; 371–386.
[4] *British Jour. Psych.*, 4, 1911, 386–456.
[5] *Psych. Bull.*, 6, 1909, 44.
[6] 'Psych. of Learn.,' 1913, 369–376.

overemphasis upon "precision." He states that if his own explanation of transfer is mysterious, that of Müller[1] is equally so. In confirmation of his conclusions, he refers to the experiments of Winch,[2] Daniel Starch,[3] Coover and Angell,[4] and Wallace Wallin,[5] which he reviews as the 'Pädagogisch wichtigste.' Nowhere does he refer to the experiment of Thorndike and Woodworth[6] which was published three years before his own, although the remarks upon special memories in the introduction of the report of his experiment have a striking similarity to the opening remarks of Thorndike and Woodworth upon the same subject. Nor does he take notice of other experiments reaching different conclusions, *e. g.*, those of James,[7] Fracker,[8] Klein,[9] Foster,[10] Ruger,[11] and Sleight,[12] all of which are reviewed in a book to which he makes more than one favorable reference.[13] It appears, therefore, that Meumann never retracted the fundamental conclusions of his experiment.

Since this experiment of Meumann forms the background of most that is said in his 'Psychology of Learning,' and of much that is said in his 'Vorlesungen,' and has elsewhere occupied a prominent place in psychological literature, and coming as it does from the hand of one of the foremost psychologists of his time, it appeared to the writer that it ought to be repeated, and that in this way many of the important criticisms that have been made upon it might be answered. I shall refer to this repetition as the test experiment which I now wish to describe.

There were two series of tests: I., memory-span tests, and, II., learning tests.

[1] *Vorlesungen*, III., 1914, p. 143 ff.
[2] *British J. of Psych.*, 2, 1908, 284–293; 3, 1910, 386–405.
[3] *J. of Ed. Psych.*, 3, 1912, 209–213.
[4] *Amer. J. of Psych.*, 18, 1907, 237–340.
[5] *J. of Ed. Psych. Mon.*, 1911.
[6] *Psych. Rev.*, 8, 1901, 247–261; 384–395.
[7] *Principles of Psych.*, 1, 1893, 667.
[8] *Psych. Rev. Mon. Suppl.*, 4, 1908, 56–102.
[9] Klein, Thorndike, 'Psych. of Learn.,' 401 ff.
[10] *Jour. of Ed. Psych.*, 2, 1911, 11–21.
[11] *Archives of Psych.*, No. 15, 1910.
[12] *British Jour. of Psych.*, 4, 1911, 386–456.
[13] Thorndike's 'Psych. of Learn.'

I. The memory-span tests consisted of:

1. Thirteen double series of consonants ranging from five to seventeen in the first test. In the second, the quantity was increased to twenty letters.

2. Sixteen double series of numbers ranging from five to twenty. In the second test, the largest quantity was seventeen.

3. Eight double series of nonsense syllables ranging from four to eleven.

4. Twelve series of disconnected one-syllable words ranging from four to fifteen.

5. Nine series of Latin-English vocabularies ranging from three to eleven pairs, tested by the Treffer method.

6. Passages from Wordsworth's 'She was a phantom of Delight,' The 'Solitary Reaper,' 'Stepping Westward,' and 'To a Highland Girl,' with the following number of words per passage in the first test: 12, 14, 16, 18, 21, 22, 24, 26 and 28. In the second test, the quantities were: 11, 14, 16, 18, 20, 22, 24, 26, and 28.

7. Eleven passages of prose from Locke's 'Essay,' ranging from twelve to thirty-four words.

The first four of the above tests were presented in an auditory form, at the rate of one member a second. The last three tests were also read aloud at a slow rate, but not at the rate of one member per second. The reproduction in all cases was written upon paper by the subject immediately after the stimulus ceased.

II. The learning tests were all presented visually and were relearned after twenty-four hours. They consisted of:

8. Three series of nonsense syllables, consisting of ten, twelve, and sixteen syllables respectively. The syllables were made from the English alphabet, and according to the conditions of Müller and Schumann.

9. Two series of meaningless visual diagrams of twelve members each, the same as those used by Ebert and Meumann.

10. Thirty Latin-English vocabularies.

11. Forty Latin-English vocabularies.

12. Two eight-line stanzas of poetry selected from Wordsworth's 'To the Daisy.'

13. A ten-line passage from Locke's 'Essay.'

The experimenter's function in the learning test was to see that the subject carried out his instructions, and to keep account of the number of repetitions made and of the time, which was taken with a stop watch. The time was measured from the beginning of the test to the end of the first correct reproduction. The same measurements were also kept for the syllable rows in the practice series. In the calculation of the data, only the times were taken into account.

The test experiment was made with fourteen adult subjects, but later reduced to thirteen, at the University of Wyoming during the summer school session of 1915. Eight of these constituted Group I., and the other six, later reduced to five, constituted Group II.

The eight subjects of Group I. did both the practice and the test series of the experiment while the six of Group II., later reduced to five, omitted the practice series but did the test series at the same time as Group I. Group II. was the control group. The two groups were divided with respect to equal learning ability as accurately as this could be determined by the experimenter's personal judgment, but, as a matter of fact, the average learning ability of Group II. was slightly superior, which makes the results for control purposes all the better.

Like the experiment of Ebert and Meumann, the test experiment had a practice series which consisted of learning and relearning twelve-syllable nonsense series, constructed as explained above. Those which were learned one day were relearned twenty-four hours later. The form of presentation was somewhat different from that in Test 8. Instead of the experimenter allowing the subject to reproduce the syllables at his own rate, he kept presenting them in a mechanical way at the rate of one per second until the learner succeeded in reproducing for the first time each syllable before it was turned. The series was then considered learned. The four methods of learning: the whole, the part, and the two mixed methods, followed by Ebert and Meumann's subjects, were also followed by the subjects of Group I. of the test experiment. The order

of the different learning methods was varied irregularly in the test experiment. Ebert and Meumann kept the quantity of learning material constant for each practice exercise so that each subject learned two series of syllables daily. But in the test experiment, the constancy was put upon the amount of time. This was limited to thirty minutes for each exercise, except when the subject failed to learn one twelve-syllable row. In that interval, each subject relearned what he had learned the day before, and then learned as much new material as possible in the space of time remaining. Four of the subjects in Group I. had two practice exercises daily, one in the morning and one in the afternoon, while the other four had only one practice exercise daily in the morning. Several of the subjects omitted Sundays and other holidays. Approximately fifteen days were spent in practice by Group I., while upon the five days immediately before and after the practice, Groups I. and II. took the tests described above. The test experiment may therefore be considered as repeating only the first practice period of Ebert and Meumann together with the tests that came before and after that period. If this part of the experiment is confirmed or refuted, the writer is of the opinion that a repetition of the remaining parts would not essentially change the character of the results.

The course of improvement in the training series is represented in the practice curves of Fig. 1 and Fig. 2. The points in the ordinates are based upon the average learning and relearning times per series for each day of practice. That the character of the improvement in learning nonsense syllables, which is memorial or ideational in function, is of the same sort as that of learning telegraphy or typewriting and of other sensori-motor functions is evident from the character of the curves, which are typical learning curves. The individual differences, however, are striking.

The relearning curves either vary about a horizontal line or else run closely parallel to the learning curves. Contrary to our expectation from the results of Ebbinghaus, Müller, and Schumann, Radosawljevitch on the relation of the amount saved to the strength of the imprinting, a high learning time

is in most cases not followed by a correspondingly low learning time, but rather the converse. But it should be said that no special light was thrown upon the causes of improvement. In the cases under observation, they appeared not to consist in lack of effort or interest, or the development of bad learning methods such as unusual pronunciation. Nor did the coöperation of the will appear to be of absolute and controlling importance, as it is stated by Meumann, for no exertion of the will prevented bad records or plateaus. See his 'Psychology of Learning,' pp. 64, 66, 74, 115, 139, 283, 287, and 361.

But whatever may be the cause of improvement, the curves of Fig. 1 and Fig. 2 show that a very great improvement was made by the subjects of the test experiment in learning nonsense syllables. The question now is: How much of this improvement was transferred to other mental functions?

The measures employed by Ebert and Meumann to determine the effect of the improvement gained in learning nonsense syllables upon the memory-span functions were the memory-span, the first excess of one third error, and the first excess of one half error. The measure of the memory-span was taken as the number of members in the largest series correctly reproduced in any given test. The amount of error was calculated as follows: "The omission or addition of a member was considered as 4/4 error, a displacement in the series of more than one position as 3/4 error, and the displacement of the same sort of exactly one position as 2/4 error and a correction as 1/4 error (*op. cit.*, p. 11)."

In the test experiment, the three above measures were used in the same way. But to increase the reliability of the tests five additional measures were used as follows:

1. The *method of average error* per quantity on series in each test calculated according to the scale of Ebert and Meumann for each series. From the results thus obtained, the average was calculated for all the series in a test together with the average deviation from the individual series.

2. The *method of retained members*, according to which the number of members correctly reproduced regardless of order and position was calculated in each series of a test. From

these results the average and the average deviation were calculated as in 1.

3. The *method of retained groups*, according to which the largest number of members reproduced that had the same

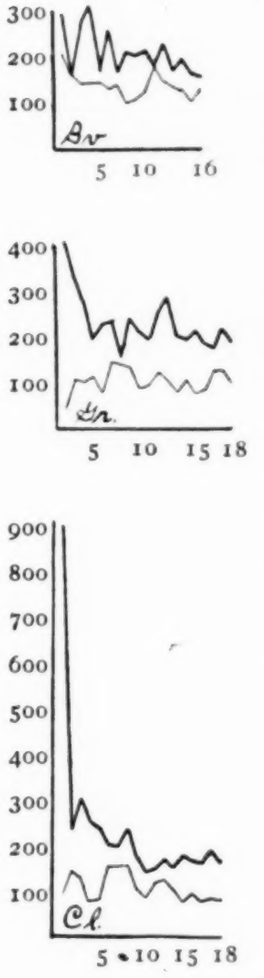

FIG. 1.

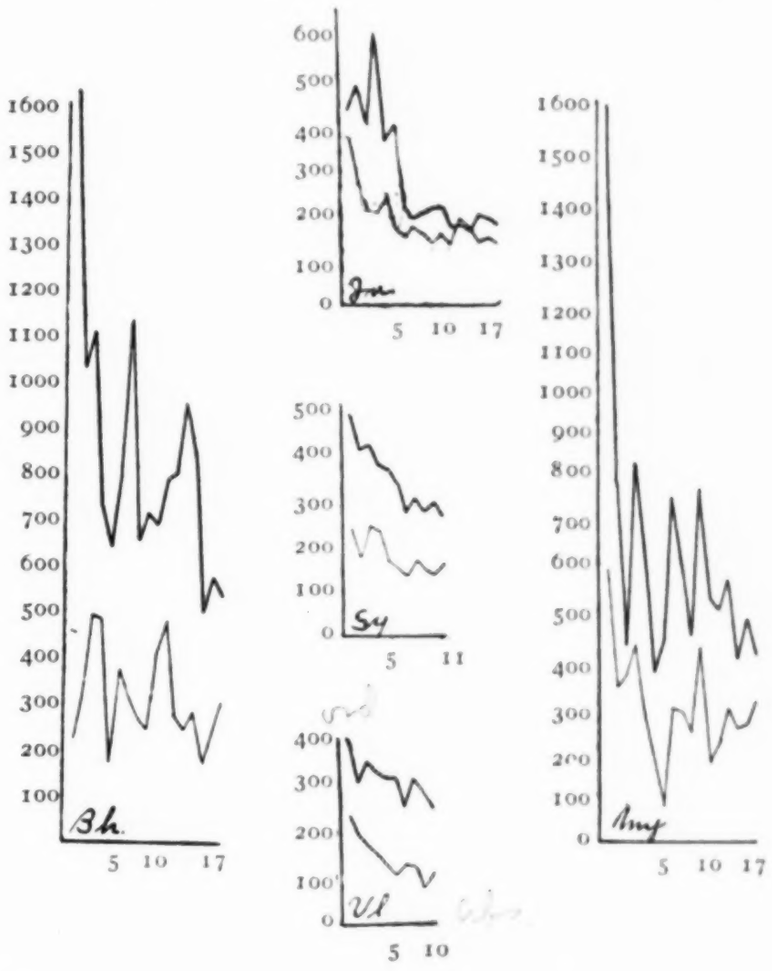

FIG. 2.

FIGS. 1 AND 2. Practice Curves Showing Course of Improvement in Learning and Relearning Twelve-syllable Rows of Nonsense Syllables. The time in seconds is plotted along the ordinates and successive practice days along the abscissas. The heavy line above represents the average learning time per series from day to day; and the light line below, the corresponding relearning time.

order of succession as those in the original series was calculated for each series in a test. The average and average deviation were determined as in 1.

4. The *method of retained positions*, according to which the number of members in the reproduction that had the

same position relative to either the first or the last end of the series was calculated. The average without the A.D. was determined as in 1.

5. The method of correlation by the Pearson Coefficient, the application of which will be described later.

In Table I., the results of the above measures except the last one are given from the various tests for Group 1 and for Group 2.

The first column in Table I., 'Highest No. Correct,' represents the memory-span measurements. The second, third, fourth, sixth, eighth, and tenth contain the measurements in the order described above. Test I. represents the tests made before the practice series and Test II. represents the tests made after the practice series. The averages opposite I. and II. are the averages, and the A.D.'s under them are the deviations, of the individual averages from the group averages. D_1 represents the differences between tests I. and II. for Group 1, while D_2 represents the corresponding differences for Group 2. A plus indicates the amount of gain in Test II. over Test I. A minus indicates a corresponding loss. D_1–D_2 the difference in gain or loss between groups 1 and 2. The A.D.'s in the vertical column after the rubrics, 'Error per Quantity,' 'Mems. Repr. per Quantity,' etc., are the deviations of the individual subjects in the test in question. The reason that the A.D.'s of 'The Error per Quantity' are so large is that the increasing the quantity of learning material beyond the memory-span increases the amount of error enormously. Thus, for example, when five letters were read the amount of error in the reproduction for most of the subjects was zero, but when a series of twenty letters were read, the amount of error for most of them was between sixteen and seventeen. This feature will become clearer when we shall discuss the influence of the quantity of learning material upon the reproduction.

From Table I. it will be seen that the value of the 28 gains of Group 1 in Test II. is destroyed by the counterbalance of 19 losses, by the fact that only two of them were more than twice as great as the A.D.'s of the averages from which they

TABLE I

LETTERS

Group	Test		Highest No. Correct	First Excess of ⅓ Error	First Excess of ⅔ Error	Error per Quantity	Aver. A.D.	Mems. Repr. per Quantity	Aver. A.D.	Size Largest Group Repr. per Quantity	Aver. A.D.	No. of Positions Correctly Repr. per Quantity
1	I	Ave.	6.75	8.12	9.12	7.48	4.10	6.53	1.37	3.25	1.35	4.47
		A.D.	1.31	.91	1.13	.52	.36	.25	.14	.47	.21	.60
	II	Ave.	6.75	7.13	8.88	9.23	4.78	7.30	1.36	3.66	1.31	4.67
		A.D.	.63	.56	.91	.51	.37	.59	.26	.48	.16	.47
		D_1	000	− .99	− .24	−1.75		+ .77		+ .41		+ .20
2	I	Ave.	8.00	9.67	10.50	7.47	4.99	7.44	1.25	3.63	1.50	4.76
		A.D.	.67	.78	1.20	.79	.32	.74	.13	.63	.19	.75
	II	Ave.	7.33	7.67	8.83	9.66	5.53	7.03	1.43	3.57	1.30	4.65
		A.D.	.44	1.67	2.16	.79	.52	.50	.22	.59	.16	1.02
		D_2	− .66	−2.00	−1.67	−2.19		− .41		− .06		− .11
		D_1–D_2	+ .66	+1.01	+1.43	+ .44		+1.18		+ .47		+ .31

NUMBERS

1	I	Ave.	8.00	9.38	10.75	7.47	4.59	8.56	1.39	4.82	1.34	5.75
		A.D.	1.50	1.87	1.75	.39	.78	1.13	.29	.84	.42	.81
	II	Ave.	8.25	9.75	10.75	5.76	4.18	8.01	1.60	4.97	1.42	6.27
		A.D.	1.31	1.44	1.25	1.60	.30	.73	.25	.65	.43	.82
		D_1	+ .25	+ .37	− .50	+1.71		− .45		+ .15		+ .52
2	I	Ave.	8.83	10.50	11.83	6.86	4.65	9.18	2.10	4.96	1.63	6.11
		A.D.	1.11	1.50	.89	.55	.57	.51	.35	.80	.11	1.04
	II	Ave.	9.00	11.00	12.17	5.38	4.08	4.48	1.55	4.83	1.62	6.31
		A.D.	1.67	1.67	1.83	.84	.52	.67	.24	.96	.13	.88
		D_2	+ .17	+ .50	+ .34	+ .48		− .70		− .13		+ .20
		D_1–D_2	+ .08	− .13	− .84	+1.23		+ .25		+ .28		+ .32

SYLLABLES

1	I	Ave.	4.37	5.38	6.13	5.40	3.08	4.15	.74	1.68	.89	2.02
		A.D.	.81	1.13	1.09	.36	.68	.48	.11	.43	.36	.64
	II	Ave.	4.54	6.50	7.13	5.55	3.44	4.74	.88	1.96	.94	2.54
		A.D.	.71	.63	.88	.24	.52	.46	.14	.41	.19	.37
		D_1	+ .17	+1.12	+1.00	− .15		+ .59		+ .28		+ .52
2	I	Ave.	5.17	6.50	7.00	5.35	3.23	4.47	.71	1.92	1.03	2.32
		A.D.	.83	.67	.67	.89	.42	.54	.10	.48	.25	.55
	II	Ave.	4.66	6.33	6.83	5.56	8.29	4.54	.83	2.10	.89	2.59
		A.D.	1.83	1.67	1.17	2.80	.16	.47	.17	.42	.22	.67
		D_2	− .51	− .17	− .17	− .21		− .07		+ .18		+ .27
		D_1–D_2	+ .68	+1.29	+1.17	+ .06		+ .66		+ .10		+ .25

WORDS

1	I	Ave.	5.13	6.25	6.50	6.39	3.65	5.34	1.19	2.84	1.09	3.36
		A.D.	.91	.75	.75	.50	.54	.50	.28	.51	.24	.73
	II	Ave.	5.00	6.00	6.00	6.47	3.45	5.59	1.04	2.79	.95	3.73
		A.D.	.67	.78	.75	.44	.45	.51	.27	.46	.17	.52
		D_1	− .13	− .25	− .50	− .08		+ .25		− .05		+ .37

TABLE I—*Continued*

WORDS—*Concluded*

Group	Test		Highest No. Correct	First Excess of ⅓ Error	First Excess of ½ Error	Error per Quantity	Aver. A.D.	Mems. Repr. per Quantity	Aver. A.D.	Size Largest Group Repr. per Quantity	Aver. A.D.	No. of Positions Correctly Repr. per Quantity
2	I	Ave.	5.67	7.17	7.33	6.57	4.11	5.42	.91	2.93	1.01	3.69
		A.D.	.56	.89	1.11	.72	.31	.41	.27	.46	.13	.57
	II	Ave.	3.17	6.50	6.50	6.95	4.04	5.00	.80	2.53	1.00	3.27
		A.D.	.28	.67	.83	.45	.56	.33	.26	.42	.09	.58
		D_2	− .50	− .67	− .83	− .38		− .42		− .40		− .42
		D_1–D_2	+ .37	+ .42	+ .33	− .53		+ .67		+ .35		+ .79
		VOCABULARIES										
1	I	Ave.	4.64	4.71	4.71	4.35	2.38	3.27	1.02	3.45	.97	3.18
		A.D.	.23	.08	.08	.52	.10	1.16	.21	.96	.48	1.04
	II	Ave.	4.29	4.30	5.14	3.67	1.96	3.87	1.10	2.60	.99	3.57
		A.D.	1.40	1.11	1.63	.97	.29	1.00	.21	.94	.43	1.44
		D_1	− .35	− .41	+ .43	+ .68		+ .60		+ .15		+ .39
2	I	Ave.	3.80	5.00	5.20	4.17	2.38	3.05	.78	2.23	.86	2.42
		A.D.	1.00	1.20	1.04	1.49	.36	.54	.31	.52	.08	.77
	II	Ave.	3.40	4.80	5.60	4.06	2.38	3.58	.74	2.31	.82	2.89
		A.D.	.88	.64	1.12	.68	.44	.40	.11	.60	.11	.80
		D_2	− .40	− .20	+ .40	− .09		+ .53		+ .08		+ .47
		D_1–D_2	+ .75	+ .21	+ .03	+ .77		+ .07		+ .07		− .08
		POETRY										
1	I	Ave.	18.55	18.28	19.13	9.21	7.87	14.29	3.24	9.88	4.08	11.76
		A.D.	2.19	1.07	1.62	3.34	1.47	1.71	.67	2.06	1.11	2.68
	II	Ave.	10.32	14.86	16.30	12.72	6.43	13.11	2.93	8.02	2.60	10.35
		A.D.	4.40	2.45	3.19	3.07	1.02	1.99	.85	2.36	.86	5.73
		D_1	−8.23	−3.42	−2.83	−3.51		− .18		−1.86		−1.41
2	I	Ave.	18.60	20.80	20.80	7.66	7.01	11.75	3.38	16.00	3.10	14.22
		A.D.	2.28	2.24	2.24	1.60	1.15	1.65	.64	1.35	.40	1.83
	II	Ave.	14.20	17.20	17.60	12.25	8.17	14.06	3.52	9.35	2.84	11.33
		A.D.	1.52	1.44	1.92	2.37	1.86	1.36	.46	1.31	.32	2.21
		D_2	−4.40	−3.40	−3.20	−4.59		+2.31		−6.65		−2.89
		D_1–D_2	−4.83	+ .18	+ .37	+1.08		−2.49		+4.79		+1.48
		PROSE										
1	I	Ave.	15.00	17.00	17.50	15.04	9.97	14.59	8.44	8.84	3.46	10.36
		A.D.	1.50	1.50	1.50	1.44	.66	1.55	.35	1.13	.86	1.52
	II	Ave.	17.00	20.00	20.50	13.07	8.09	16.43	3.86	10.70	3.50	12.92
		A.D.	3.50	2.50	2.15	2.15	.79	1.55	.34	1.99	.82	2.06
		D_1	+2.00	+3.00	+3.00	1.97		1.84		1.86		2.56
2	I	Ave.	16.66	16.33	21.33	14.70	1.24	14.90	3.08	9.40	3.48	10.48
		A.D.	1.88	2.33	3.33	2.56	.74	3.12	.47	1.75	.18	1.46
	II	Ave.	20.00	22.00	23.20	12.45	8.66	17.23	3.71	11.05	3.54	11.55
		A.D.	3.20	.00	1.92	3.23	1.52	2.05	.95	2.23	.54	1.68
		D_1	+3.34	+5.27	+1.87	+1.25		2.33		1.65		1.07
		D_1–D_2	−1.84	−2.67	+1.13	+ .72		+ .49		+ .21		+ .49

were computed, and by the 19 gains made by Group 2 in the same tests. When the difference in gain and loss between the groups is considered, we discover that Group 1 had thirty-nine gains over Group 2, and against these ten losses must be counted. (This situation in itself apparently indicates that Group 1 increased its memory-span somewhat by the training in nonsense syllables, but, when the character of the gains and losses which make up this situation are considered, the facts of Table II. can only mean that the *special training had no effect in either increasing or decreasing the memory-span* for the various materials with which the tests were made. The important condition of this conclusion is that neither Group 1 nor Group 2 made any significant gains or losses in Test II. as compared with Test I.)

Table II. shows the average performances of Groups 1 and 2 in the learning tests. *L* = Learning; *RL* = Relearning; ′ = minutes and ″ = seconds. The other symbols are the same as in Table I. The A.D.'s represent the deviation of the individuals of the group from the group average.

Table II. gives the measurements of the average performances of Groups 1 and 2 in the learning tests. Group 1 made three gains and three losses in the learning functions. Only one of the gains is greater than the A.D. of the average of either Test I. or Test II., and the other two are less. Group 2 made three gains and three losses in the learning functions, and all of them are less than the A.D.'s of the averages of either Test I. or Test II. (This can only mean that there is clearly no *general learning ability that is improved by the special* training in nonsense syllables. Although no general learning ability is improved by the special training in nonsense syllables, it is still possible that this training might have improved some special learning function.) In the test with nonsense syllables, Group 1 gained 6′ 47.3″ over Group 2. In the diagrams, Group 1 gained 5′ 12.9″ over Group 2. These gains are much greater than those gained in any other functions. When these gains are not considered with reference to the A.D.'s, they indicate that the special training did improve the ability to learn nonsense syllables and nonsense

TABLE II

	Test	′	L″	A.D.″		RL″		A.D.″	
Syllables									
1.	I.	12	44.2	6	11.9	4	47.0	1	53.1
	II.	7	34.7	4	5.6	3	5.6	1	10.4
	D_1.	+ 5	95			+1	41.4		
2.	I.	7	7.2	2	11.6	3	52.4	1	24.8
	II.	8	46.5	3	32.6	3	29.0	0	52.9
	D_2.	− 1	39.3			+0	23.4		
	D_1–D_2.	+ 6	47.3			+1	18.0		
Diagrams									
1.	I.	17	18.2	6	10.3	5	57.5	3	12.4
	II.	9	36.3	4	37.3	3	38.9	0	59.6
	D_1.	+ 7	31.9			+2	18.6		
2.	I.	12	11.6	3	8.4	5	49.7	0	50.4
	II.	9	52.6	2	55.5	3	20.4	1	4.2
	D_2.	+ 2	19.0			+2	29.3		
	D_1–D_2.	+ 5	12.9			−0	10.7		
30 Vocabularies									
1.	I.	18	34.8	5	10.7	7	5.6	4	53.3
	II.	19	58.5	6	41.9	5	42.0	2	43.3
	D_1.	− 1	18.7			+1	23.6		
2.	I.	23	29.8	7	50.5	8	29.6	3	55.5
	II.	18	38.3	6	37	6	34.5	1	42.8
	D_2.	+ 4	50.5			+1	55.1		
	D_1–D_2.	− 6	9.2			+	31.5		
Poetry									
		21	29.1	13	45.5	3	45.7	1	27.5
		19	16.4	10	49.0	3	29.4	1	32.4
		2	12.7			+0	16.3		
		18	44.5	6	45.6	4	29.7	1	40.2
		18	49.9	6	6.8	4	52.4	2	32.8
		− 0	5.4			−0	22.7		
		+ 2	18.1			+0	39.0		
Prose									
		23	22.7	8	51.8	7	35.5	5	24.0
		23	27.7	10	27.7	5	18.8	3	10.7
		− 5	05.0			+2	16.7		
		29	48.3	13	54.1	5	38.5	1	0.6
		27	43.0	11	20.4	8	2.3	3	37.7
		+ 2	05.3			−2	23.8		
		− 2	10.3			+4	40.5		
40 Vocabularies									
		23	59.0	9	34.9	7	11.8	4	17.9
		27	34.7	7	30.0	8	40.8	3	30.7
		− 3	35.7			−1	29.0		
		31	34.6	13	22.1	7	51.8	4	6.4
		35	55.6	15	19.4	6	36.2	1	46.1
		− 4	11.0			+4	15.6		
		+ 0	35.3	− 2	44.6				

visual diagrams. We must bear in mind, however, that Group 1 might gain more in a given test than Group 2, without the greater amount of improvement being due to special training.

In the relearning functions, Group 1 has five gains and one loss, but in all cases, at least one of the A.D.'s of the average from which these gains or losses were computed, is greater than the gain or loss in question. Group 2 has four gains and two losses. Only one of the gains is greater than the A.D. of the group averages from which it was computed. These facts undoubtedly mean not only that there is no general relearning ability which is improved by special training in learning nonsense syllables, but also that the efficiency of no special learning function is apparently affected by the special training.

The measurements given in Tables I. and II. may be supplemented by the measurement of the Pearson coefficient of correlation. If an individual's memory-span and his learning ability are allied, or if both are allied to some general ability, and if one of these factors is improved by a special training series, there ought to be a positive correlation between the performances in the training series and the performances in the test series, and the correlation between the training series and the second test series ought to be higher than that between the training series and the first test series.

Since there were seven measurements in the memory-span tests, it is possible that one of these is more reliable than another. The rank of the individuals in each test in a given measurement was correlated with their rank in the same test in each of the other measurements. For example, the memory-span for letters was correlated with the amount of error in letters, with the average size of the largest group reproduced in letters, and with the average number of positions correctly reproduced in letters. The 'First Excess of 1/3 Error' and 'The First Excess of 1/2 Error' were not considered in the correlation. Only the performances of Group 1 in Test I. were considered. Tables III. and III. *A* show the results of these computations. In Table IV. are shown the correlations

TABLE III

SHOWING THE CORRELATION, R, BETWEEN THE DIFFERENT MEASURES OF THE PERFORMANCES IN THE MEMORY-SPAN TESTS, GROUP I, TEST II.

$$R = 1 - \frac{6\Sigma g}{n - 1}$$

Tests

Measures / Correlation, R, of Memory-span with	Letters	Numbers	Syllables	Words	Vocabularies	Poetry	Prose	Ave.
Amount of error	.05	.48	.14	.48	.91	.52	.48	.44
No. of members	.24	.29	.53	.48	1.00	.22	.22	.43
Size of groups	.64	.53	.48	.48	.91	.48	.68	.60
Positions	.43	.48	.38	.22	.91	.31	.48	.46
Amount of error with no. of members	.14	.81	.34	.34	.91	.81	.34	.53
Size of groups	−.14	.53	.15	.71	1.00	.91	.71	.59
Positions	.24	.72	.24	.64	1.00	.81	.67	.62
No. of members with size of groups	.23	.48	.34	.43	.91	.72	.53	.52
Positions	.34	.62	.34	.14	.91	.81	.57	.53
Size of groups with positions	.22	.48	.81	.53	1.00	.91	.67	.66
Ave.	.24	.54	.38	.45	.95	.65	.54	

between each test in a given measure and the same test in other measures. In Table III. *A* the averages of these seven correlations between any two measures are shown. The general averages given at the bottom of Table III. *A* may be considered as giving the relative reliability of the five measures

TABLE III*A*

SHOWING THE AVERAGE CORRELATIONS BETWEEN THE DIFFERENT MEMORY-SPAN MEASURES AS COMPUTED FROM ALL THE TESTS, THE INDIVIDUAL CORRELATIONS OF TABLE III

	Memory-span	Amount of Error	No. of Members	Size of Groups	Position
Memory-span		.44	.43	.60	.46
Amount of error	.44		.53	.59	.62
No. of members	.43	.53		.52	.53
Size of groups	.60	.59	.52		.66
Positions	.46	.62	.53	.66	
Ave.	.48	.55	.50	.59	.57

that were correlated. If so, the Size-of-the-Groups measure is the most reliable, having a correlation of .59; and the Memory-span measure is the least reliable, having a correlation of .48. In the same way, the averages at the bottom of

Table III. may be considered as giving the relative reliability of the seven memory-span tests, and, if so, the test with vocabularies is the most reliable, having a correlation of .95; and the test with letters is the least reliable, having a correlation of .24.

Having determined the reliability of the different measurements, the next step was to measure the spread of improvement from the training series by the Pearson coefficient of correlation. The rank of the individuals in the training series was determined by the average learning time of all the syllable

TABLE IV

SHOWING THE CORRELATIONS, r, BETWEEN THE AVERAGE LEARNING ABILITY, A.L.A., IN THE PRACTICE SERIES WITH EACH OF THE TESTS BEFORE (I) AND AFTER (II) THE PRACTICE

$$r = \frac{\Sigma x \cdot y}{\sigma x \sigma y}$$

A. Memory-Span Tests, with Memory-Span Measure

	Letters		Numbers		Syllables		Words		Vocabs.		Poetry		Prose	
Test	I	II	I	II	I	II	I	II	I	II	I	II	I	II
A.L.A.	.69	.32	.65	.47	.23	.86	.52	.15	.66	.68	.44	.82	.30	.58

B. Memory-Span Tests, with Size of Group Measure

	Letters		Numbers		Syllables		Words		Vocabs.		Poetry		Prose	
Test	I	II	I	II	I	II	I	II	I	II	I	II	I	II
A.L.A.	.60	.65	.59	.60	.34	.38	.20	−.04	.51	.74	.35	.98	−.60	.66

C. Learning Tests

	12 Syllables		Diagrams A		30 Vocabs.		40 Vocabs.		Poetry		Prose			
Test	I	II	I	II	I	II	I	II	I	II	I	II		
A.L.A.	.91	.95	.26	−.47	.95	.77	.93	.81	.84	.77	.83	.94		

rows learned in that series. Their rank in the memory-span tests was determined both by the memory-span measure, since it was the principal measure of Ebert and Meumann, and the Size-of-the-Groups measure, since it had the greatest reliability. The rank of the individuals in the learning tests was determined by the learning time since it was the only measure used. Table IV. gives the results.

The facts of Table IV. *A* show a rather high positive corre-

lation between the average learning ability for nonsense syllables and the memory-span for each of the first series of the tests. These correlations are improved in the second series of tests in the case of nonsense syllables, vocabularies, poetry, and prose; and they are reduced for the other tests.

The correlations between the average learning ability and the memory-span tests, as determined by the Size-of-the-Groups measure, which are given in Table V. *B*, compare favorably with those of Table V. *A*. There are two cases, however,

TABLE V

THE RESULTS ARE IN MINUTES, SECONDS, AND PERCENTAGES

A. Ebert and Meumann

	(1)	(2)	D_1	Per Cent. 1
Learning	3′ 22″	2′ 1.6″	1′ 20.4″	59.7
Relearning	1′ 24.2″	1′ 6.0″	0′ 18.2″	77.9

B. Test Experiment

Learning	8′ 27.8″	5′ 14.7″	3′ 13.1″	63.7
Relearning	3′ 40.1″	2′ 34.9″	1′ 5.7″	74.8

where the relationship between Tests I. and II. changes direction, namely, in the tests with letters and numbers.

The correlations of the average learning ability in the training series with the first series of learning tests are on the whole higher than those with the memory-span tests, which indicates a closer relationship between these functions. These correlations are increased in the second series of learning tests for nonsense syllables and for poetry. The rank in the case of the first was determined by the twelve-syllable test series alone, since it had the same quantity as those in the test series. The most significant gain is made with this test, thus confirming the result of Table II., but in case of the diagrams, a negative correlation of .47 appears in the second test as against the positive correlation of .26 in the first test. This drop of .73 means that the individuals who did the worst in the training series did the best in the second test with the diagrams. But as measured by the learning time in Table II., more gain was made in this test than in any other one. If

the negative correlation has any significance, it certainly means that the improvement as measured by the time is not due to the training series but to the discovery of some special trick by which these nonsense visual diagrams were learned with ease. Where the measurements of Tables I. and II. contradict those of Table IV., it would be hazardous to express an opinion in regard to the transfer of training. Where they do not contradict each other, it may be safe to do so. In the learning tests, there is only one such case that indicates a transfer effect from the training series, namely, the test with nonsense syllables. In the memory-span tests, there are two cases, if the gains of Group 1 over Group 2 are taken in an absolute sense without reference to the A.D.'s, namely the memory-span for nonsense syllables and the memory-span for vocabularies. But the only safe conclusion appears to be that improvement in learning nonsense syllables by one method improves the ability to learn nonsense syllables by another method. In any case, the spread of improvement from a special function is not general, but it is very specialized and affects only such other special functions that are very similar to the one specially trained. The result of the Test Experiment in regard to the transfer of training is that the findings of Ebert and Meumann and the conclusions based upon them, that there is a general memorial function and that special training in one function improves the memory in general, are not confirmed. Nor is their theory confirmed that transfer from one function to another is in proportion as the functions are allied. If the correlations in Table IV. are measures of the relations between the learning ability for nonsense syllables and the special functions experimented upon in the test series, they mean that the specially trained function is closely related to practically all of the functions measured in the test series, and that the relationship of this function to the learning functions is closer than to the memory-span functions. Either the learning and memory-span functions condition each other or both sorts are conditioned by some unknown general ability. Excepting the diagrams, the specially trained function has a relationship with all of the learning functions of the learning

tests above .75, but there is evidence of transfer to only one of these and evidence of interference with two of them. The probable meaning of this is that the conditions of improvement and of transfer are peculiar to the function especially trained and do not apply to any intrinsic relationship between the abilities of the various mental functions. What these peculiar factors are is difficult to say, but the writer has some grounds for believing that they are the special associative connections developed in the course of practice upon a given material, and in proportion as these peculiar associative connections can be applied to other materials, there is a transfer of improvement.

The next question is whether the results of Ebert and Meumann and the results of the test experiment are comparable. Did the subjects of the test experiment have as much to transfer as those of Ebert and Meumann: that is, did they improve as much? If they did not, the results of the test experiment could not be considered as modifying the results of Ebert and Meumann. We may compare the amounts of improvement attained in the two experiments by the following methods: 1. We may calculate the average learning time, (1) of the first two-syllable rows learned by each method in the training series, the same (2) for the last two by each method, and then take the ratio, per cent. 1, of the last to the first as the measure of the amount of improvement. 2. We may take the absolute difference, D_1, between these two learning times as the measure of the amount of improvement. 3. We may take the ratio, per cent. 3, of the average learning time of the last two days, (3) of the practice period to that, (4) of the first two days of the practice period. 4. We may take the absolute difference, D_3, between these two as the measure of the amount of improvement. 5. We may take all of these together. The writer has calculated all these measures from the data of the two experiments in question. The results are given in Table V. Table V. *A* shows the group averages for Ebert and Meumann's Subjects while Table V. *B* shows them for Group 1 of the test experiment.

It will be noticed that the amount of improvement as measured by the ratio of the last two of each method to the

first two of each method in the practice period of Ebert and Meumann's subjects reduced their learning time 4 per cent. more than those in the Test Experiment, 59.7 as against 68.7; but as measured by the ratio of the average learning time of the last two days of practice to the first two days of practice, they improved 10.9 per cent. less than those in the test experiment, 60.4 as against 49.5. The reason for this difference is that Ebert and Meumann's subjects learned their series as regards method in a fixed order, so that each subject covered the four methods used by them every two days. This was not the case in the test experiment in which three or four days were often required to learn two series by each method. The ratios of the last two days of the practice period to that of the first two days should, therefore, be regarded as more reliable, and, according to this measure, the subjects of the test experiment improved more than did those of Ebert and Meumann in their first practice period. If the absolute difference between the learning times between the first two by

TABLE VI

A. Ebert and Meumann

	(4)	(5)	D_3	Per Cent. 3
Learning	3′ 50.3″	2′ 1.6″	1′ 20.4″	60.4
Relearning	1′ 19.2″	1′ 6″	0′ 13.2″	82.0

B. Test Experiment

Learning	10′ 17.8″	4′ 47.9″	5′ 29.9″	49.5
Relearning	3′ 55.0″	2′ 32.0″	1′ 23.0″	66.8

each method and the last two by each method, or if the difference between the learning times of the first two days of practice and the last two days of practice be regarded as the measure of the amount of improvement, then the subjects of the test experiment improved more than twice those of Ebert and Meumann. If the quantity of material learned is important for the question at hand, Ebert and Meumann's subjects each learned thirty-two series while each of those in the test experiment learned forty-two series in the average. And with respect to time, it appears that the subjects of the test experi-

ment spent more than twice as much time in practice as those of Ebert and Meumann. The learning times for the latter were obtained by multiplying the number of repetitions by twelve seconds, the duration of one rotation of the drum from which the series were read. It appears, however, that Ebert and Meumann stopped their drum while the learner made a trial reproduction, and then started it again from the syllable on which he first halted. Such a repetition would require more than twelve seconds, thus causing the above method of calculating the learning times to be too low. However, the conclusion from the discussion is that with regard to the amount of improvement and the condition of its transfer, the test experiment is fully comparable with the one of Ebert and Meumann. Not only did its subjects spend more time in practice and learn a greater quantity of material, but they also attained a greater amount of improvement both in learning and relearning than did the subjects of Ebert and Meumann in their first practice period. Other things being equal, the subjects of the test experiment should have shown more spread of improvement than those of Ebert and Meumann, but the former, unlike the latter, instead of showing that improvement n one function raises the efficiency of every other function to a large extent, showed that it principally raises the efficiency of the one specially trained and that of other special functions to a small extent, provided they are similar.

Just what the explanation of Ebert and Meumann's results must be in the light of this conclusion is difficult to say. But certain experiments of Swift, Schuyler, Thorndike, and Book have shown that improvement in relearning after several years of no practice is much more rapid than in the original learning. In view of the extraordinary amount of memorizing done in the German gymnasium, I believe that the improvement of Ebert and Meumann's subjects in their tests must be interpreted as a rapid re-acquisition of their former memorial skill learned in the gymnasium.

The second problem discussed by Ebert and Meumann was the economy of learning. In this connection, the series

of nonsense syllables of the practice period were learned by four methods as follows: *A*. This was the whole method. The series were repeated without pause from beginning to end until the first correct reproduction was reached. *B*. The part method, in which the subject first learned the first half of the series, then the second half, and finally the two halves together. *C*. The same as the *A* method, except that the subject paused one second at the end of the fourth syllable and one second at the end of the eighth syllable. *D*. The same as the *A* method except that the subject paused two seconds at the end of the sixth syllable. The essential difference between these two methods is the size of the units in which the respective series are learned. An *A* series is learned as a unit of twelve. A *B* series is first learned in units of six, and then the two units are learned together. A *C* series is learned together from the beginning but in units of four. And a *D* series is learned together from the beginning but in units of six. These four learning methods were also followed in the test experiment, but instead of the order of succession being varied in a fixed way, so that each subject learned an equal number by each, as with Ebert and Meumann, the order was varied irregularly so that in some cases a subject learned twice as many by one method as by another. The purpose of this was to discover whether the amount of practice would be a factor in determining the most economical method of learning. The group averages on the economy of learning, as the writer has calculated them from the data of both experiments, are given in Tables VII. and VIII.

TABLE VII

(EBERT AND MEUMANN)

Method	*A*		*B*		*C*		*D*	
S's.	L.	Rl.	L.	Rl.	L.	Rl.	L.	Rl.
Ave's	10.5	4.4	9.2	4.7	7.9	4.1	9.4	4.4
L. + Rl. of ave's		14.9		13.9		12.0		13.9
Rl. after 24 hrs.		56.9		47.5		49.7		63.4

Commenting upon these results, Ebert and Meumann say that the *C* and *D* methods are the quickest in learning and

produce a retention of average security. The *D* method is relatively quick in learning but uncertain in retention. The *A* method is the slowest in learning but has the greatest security in retention and certainty in reproduction. The *C* method is by far the best in learning. The *A* method is equally the best in retention, and, as regards the speed of learning, the *B* method lies nearer to the *C* and *D* methods than to the *A* method. In general these statements are confirmed in the test experiment. The *C* method is the quickest and the *A* method is the slowest in learning. As measured by the Saving

TABLE VIII

(TEST EXPERIMENT)

Times in Minutes and Seconds

Method	*A*				*B*				*C*				*D*			
S's.	L.		Rl.		L.		Rl.		L.		Rl.		L.		Rl.	
Ave's	7	15.1	3	7.5	6	18.8	3	19.	5	49.9	2	55.2	6	48.5	2	29.4
L. + Rl.			10	22.6			9	37.8			8	45.1			9	17.1
Per cent. saved in Rl. after 24 hrs.				59.				48.3				48.3				51.7

Method, the *B* method has the poorest retention, but the *D*, and not the *A* method, has the best retention. With respect to the speed of learning, the order from the slowest to the quickest of the different methods in both experiments is the same, namely, *A*, *D*, *B*, *C*. In relearning, the order for Ebert and Meumann is *B*, *D*, *A*, *C*; and in the test experiment, it is *B*, *A*, *C*, *D*.

An important question with reference to the most economical method of learning is how to measure economy in learning. Upon this matter, Meumann says: "That method of learning is most economical which secures a particular memorial effect or attains a particular memorial purpose in the shortest time, with the least number of repetitions, and with the minimum degree of fatigue; and this method may be regarded as the most economical only with reference to this memorial effect and this memorial purpose. Of these three determinations, the learning time measures the economy of time; while the economy of energy is measured by the number

of repetitions, and in less precise form, by the amount of fatigue. No other accurate means of measurement is at our disposal.[1] Meumann puts much emphasis upon the particular memorial effect desired. One method may be most economical for immediate retention, a second for temporary retention, a third for permanent retention. The quality and quantity of the material may also determine the most economical method of learning. And the character of the associations as well as the character of the reproduction, its fidelity, rapidity, and completeness must also be considered. With all of these factors, the economy of learning becomes an extremely complex problem, and much experimentation will be needed before all of these factors can be evaluated. But it is difficult to see why the time element would not be sufficient measure of the economy of anyone of these factors, especially if economy is to be measured in terms of efficiency. For example, if one method is quickest in learning and another is the quickest in relearning, and we wish to know which is the quickest in respect to both, we simply consider that here are two pieces of work to be done, learning and relearning, and that method by which both may be done in the least time is the most economical.

In Table VII. the average number of readings required by Ebert and Meumann's subjects for both learning and relearning by each method, have been added. In Table VIII., the corresponding times appear for the subjects of the test experiment. According to this measure, the *C* method is the most economical in both experiments, a fact contrary to the conclusions in the investigation of Steffens[2] and Pentschew[3], both of whom made out strong cases for the *A*, or whole, method.

In the test experiment an effort was made to evaluate two factors that might influence the most economical method of learning, namely, practice and the memory-span. The subjects did not learn an equal number by each method. This afforded an opportunity for learning the effect of practice upon the economy of learning. The effect in question was

[1] 'Psychology of Learning,' p. 373.

[2] *Zeit. f. Psychol.* **22**, pp. 321, 465.

[3] *Arch. f. d. ges. Psychol.* 1903, **1**, p. 417.

determined by calculating the Pearson coefficient of correlation. This was a difficult matter, but space permits only to say that it resulted in +.14, a figure indicating some relationship but not enough to be of importance.

To determine the correlation between the memory-span and the size of the unit into which the different learning methods naturally divided the syllable rows, the average memory-span of each subject in all of the memory-span tests, *i. e.*, those before practice were calculated. Each of these averages was then arrayed with the size of the unit in which the subject learned the quickest as in Table IX.

TABLE IX

S's.	Ave. Memory-span	Most Economical Unit
Nl.	10.57	6
Cl.	10.30	4
Sy.	10	12
Gr.	9.85	12
Jn.	7.95	12
Bv.	7.20	6
My.	6.93	4
Bh.	6.76	4
Ave.	8.67	7.5

$r = .46$.

The Pearson coefficient of correlation as calculated from these averages is +.46, which shows that the size of the units in which the subjects learned is a much more important factor in the economy of learning than the amount of practice, and that it is a factor to be reckoned with in connection with this problem. The importance of this factor is still more evident when we compare the memory-span of Group 1 in nonsense syllables with the size of the most economical learning unit, as in Table X.

TABLE X

Method	Ave. L. T.	Size of Unit in Each Method	Ave. Mem. Span of Group 1 in Syllables
A	7′ 15.1″	12	4.37 Test I
B	6′ 48.5″	6	4.54 Test II
D	6′ 18.8″	6	
C	5′ 49.9″	4	Ave. 4.45
Ave.	6′ 30.6″	7	

$r = .91$.

From Table X. it will be seen that the most economical unit, 4, lies very near the average memory-span of Group 1, 4.45. The units of the *B* and *D* methods, which are 6, lie 1.15 farther away from the group memory-span. The unit of the *A* method, which is the slowest of all, lies 7.15 away from the group memory-span. The relationship between the quickest learning method of Group 1 and its most economical learning unit is beautifully expressed by the Pearson coefficient, which is +.91. The correlation between the optimum method and the optimum unit of the group is thus .91, while that corresponding to the individuals is .46. It is probable that the latter would approach the former if a greater number of individuals were compared, and, if so, we must consider the size of the unit in which an individual learns as the principal determinant of his optimum method. The size of this unit is conditioned by the learner's memory-span. Under these conditions, we could not advise the whole method as such, or the part method as such, or the modified whole method as such, as the optimum method, but rather that an individual should learn in such units that lie near his memory-span for the material in question.

It is possible, however, that the character of the associations may have something to do with the optimum method. It will be noticed from Table X. that the learning time for the *B* method is shorter than for the *D* method, but its relearning time is much greater, so that the per cent. saved in the *D* method is 63.4 as against 47.5 in the *B* method. Considering both the learning and the relearning times, the *D* method is more economical than the *B* method. In the *B* method, the halves of the series were first learned separately and then together, while in the *D* method the halves were learned together from the beginning. My. and Bh. reported that learning the halves of the *B* method together after having learned them separately required a new organization of the entire series. They could not put the two halves together mechanically, but had to break up the associations formed with the halves and build new ones, thus causing interference between the two sets of associations. Such a condition would

easily explain the small saving in the *B* method, for the interference would naturally weaken both sets of associations. On the same grounds, the great saving in the *D* method may be explained as being due to the absence of such conflicting associations. But the difference in the saving between the *A* and *C* methods is better explained as due to the difference in imprinting. So far as this experiment goes, the advantage of the *C* and *D* methods is due to the fact that they do not form conflicting associations but yet adjust the material to the learner's memory-span. This would suggest that a learner should generally follow a modified whole method with pauses in between subdivisions the size of which is determined by the memory-span for the material in question.

The third problem which comes up in this experiment but which was not considered by Ebert and Meumann is the influence of quantity of learning material upon the memory-span. The memory-span tests varied so widely in quantity that they afforded an excellent opportunity for studying the effect of the quantity of the stimulus upon the quality of the immediate reproduction. A study of this sort is not only interesting for its own sake, but also for the light it throws upon the problem of the economy of learning.

The quality of the reproduction was measured in four ways: (1) the average amount of error; (2) the average number of members reproduced; (3) the average size of the largest groups reproduced; and (4) the character of the position reproduced. The nature of the first three measures was explained in the first part of this report, but a little more should be said about the positions. These were divided into three classes: (1) those that were correctly reproduced with respect to their relative positions to either end of the series; (2) those that were misplaced; and (3) those that were omitted.

The measurements of the subjects' performances in a given quantity in a given test were all placed in vertical columns, and then the averages of the measurements added vertically for both Tests I. and II. irrespective of the group divisions were calculated in order to show the effect of the quantity of the stimulus upon the amount of error. The A.D.'s of the

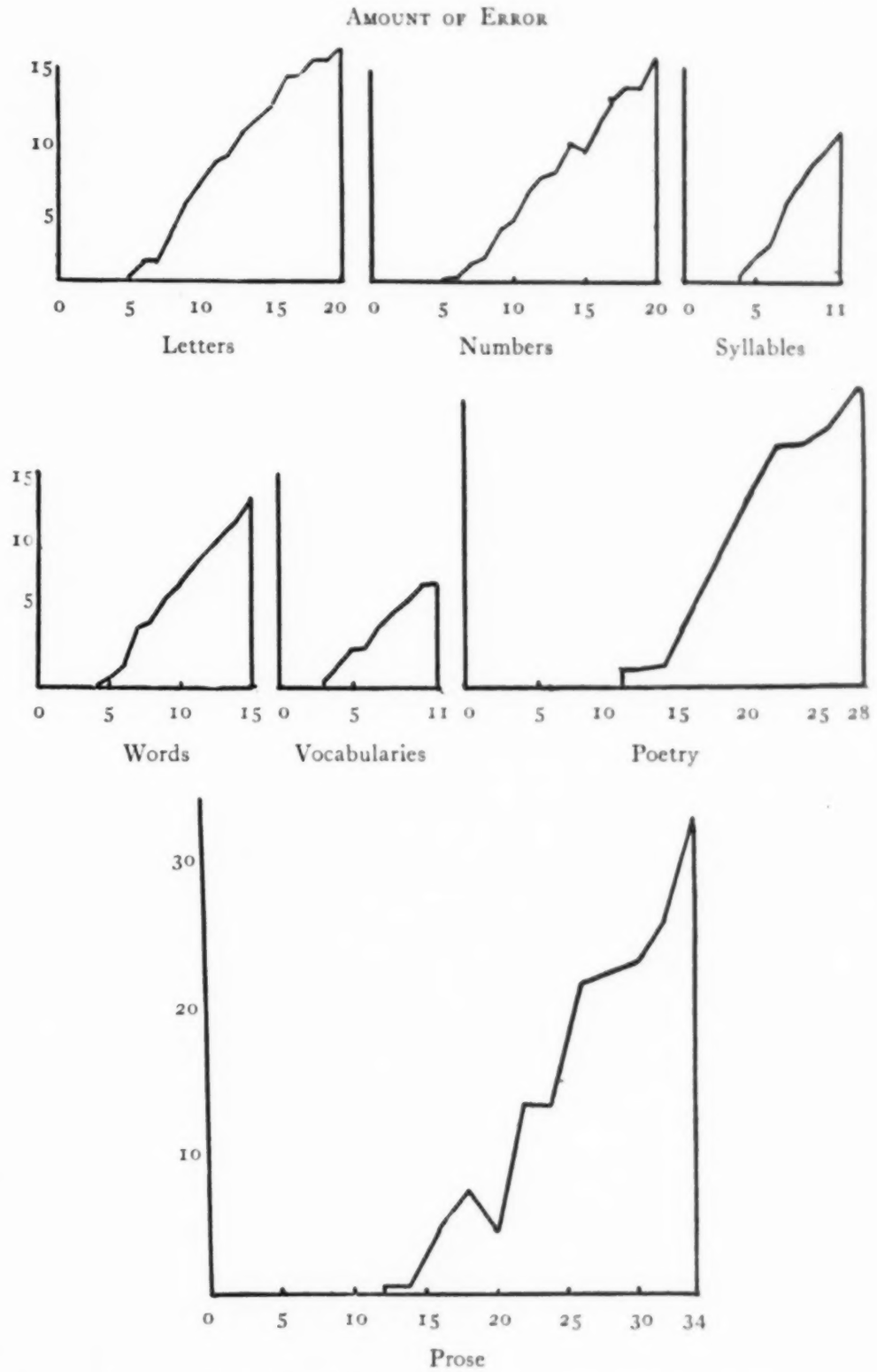

FIG. 3. Curves showing effect of quantity upon amount of error.
Ordinates = Ave. no. errors.
Abscissas = Ave. no. members in series.

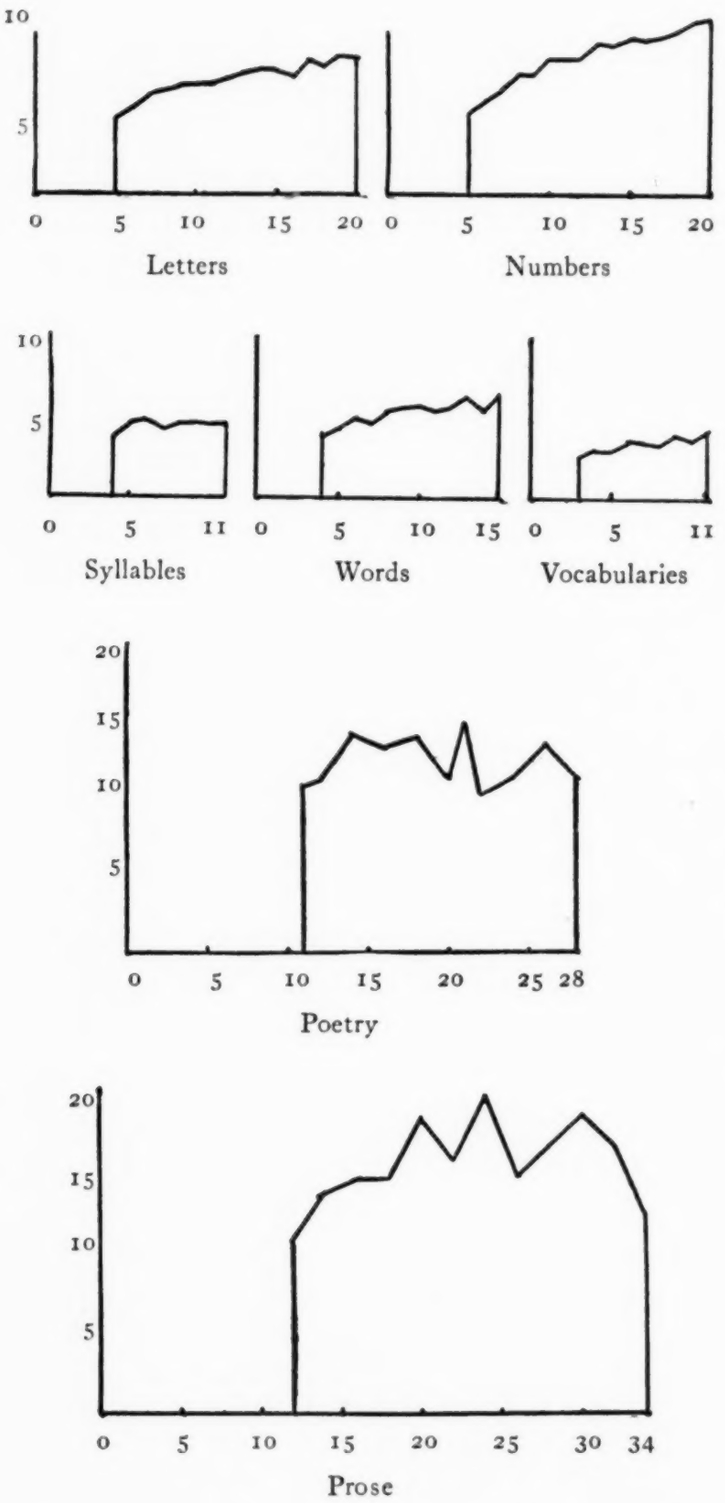

FIG. 4. Curves showing effect of quantity on number of members reproduced.
Ordinates = Ave. no. of members reproduced.
Abscissas = Ave. no. of members in series.

averages were also calculated. The averages in the tests with letters, numbers, and syllables are each based upon sixty measurements; and in those with vocabularies, words, poetry, and prose they are each based upon thirty measurements. The results with reference to the amount of error are presented graphically in Fig. 3. It will be noticed that the amount of error increases directly as the quantity of stimulus.

The results with reference to the number of members are presented graphically in Fig. 4. Perhaps the effect of quantity upon the number of members could be expressed by saying that increasing the stimulus beyond a quantum a little larger than the memory-span increases the number of members reproduced by a fraction whose value approaches zero. In fact, it is not very incorrect to say that beyond such a point the number of members reproduced is not increased. For example, in the case of syllables increasing the number of members beyond six does not increase the size of the reproduction beyond the average memory-span, 4.45. In case of prose, the number of members is not increased materially for sentences having more than 20 words, the average memory-span being a fraction over seventeen.

Fig. 5. shows the effect of quantity upon the size of the groups correctly reproduced. It will be noticed that increasing the quantity of the stimulus beyond the memory-span decreases the size of the groups correctly reproduced. This is more noticeably true of meaningful material than of nonsense material. For example, when a sentence of twenty words is read, the average size of the largest groups reproduced was 14.43 words. When a sentence of thirty-four words was read the average size of the largest group was only 4.75. When we consider that the ability to comprehend long spoken or written sentences depends not upon the ability to get a word here and there, but upon our ability to get a connection of all the words, we see the pedagogical importance of the above relationship between quantity and an immediate reproduction that is held responsible for order. This is all the more significant because most things that are read and heard are given only one repetition.

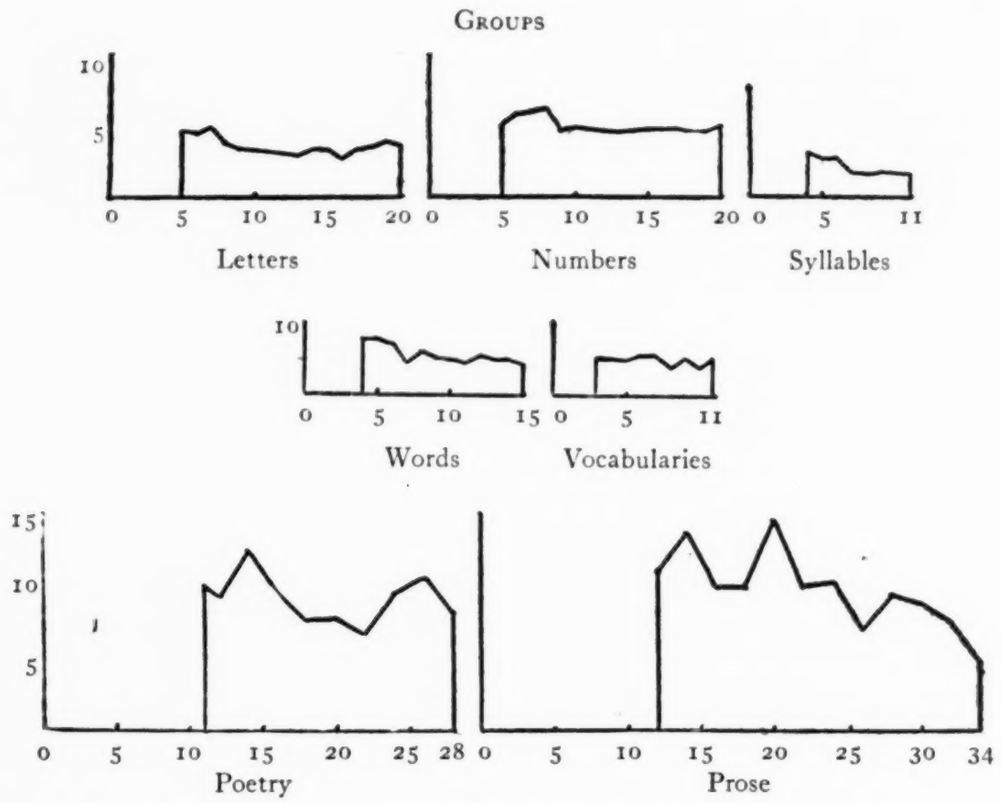

Fig. 5. Curves showing effect of quantity on size of groups.
Ordinates = Ave. size of groups reproduced.
Abscissas = No. of members in series.

The effect of quantity upon the character of the positions reproduced and omitted is shown in Fig. 6 which depicts those from the test with disconnected words. The curves here are typical of all the tests. They may be described verbally as follows: If in coördinates, the frequency of positions reproduced is plotted along the ordinates and the successive positions are placed along the abscissa line, the distributions of the positions are as follows: First, for those correctly reproduced: When the stimulus is less than the memory-span, their distribution is a rectangle. As the stimulus increases beyond this point, the roof of the rectangle begins to fall, first in the middle and then along the entire line, and, as it does this, the sides of the rectangle become a little shorter, but yet remain quite high even for large quantities, the side standing for the end of the series being higher than that for the first of the series. The roof of the rectangle thus tends to form a curve

which has the reverse shape of the probability curve. Second, the positions omitted: The curve for this omission is just the reverse of the one described. The distribution for quantities

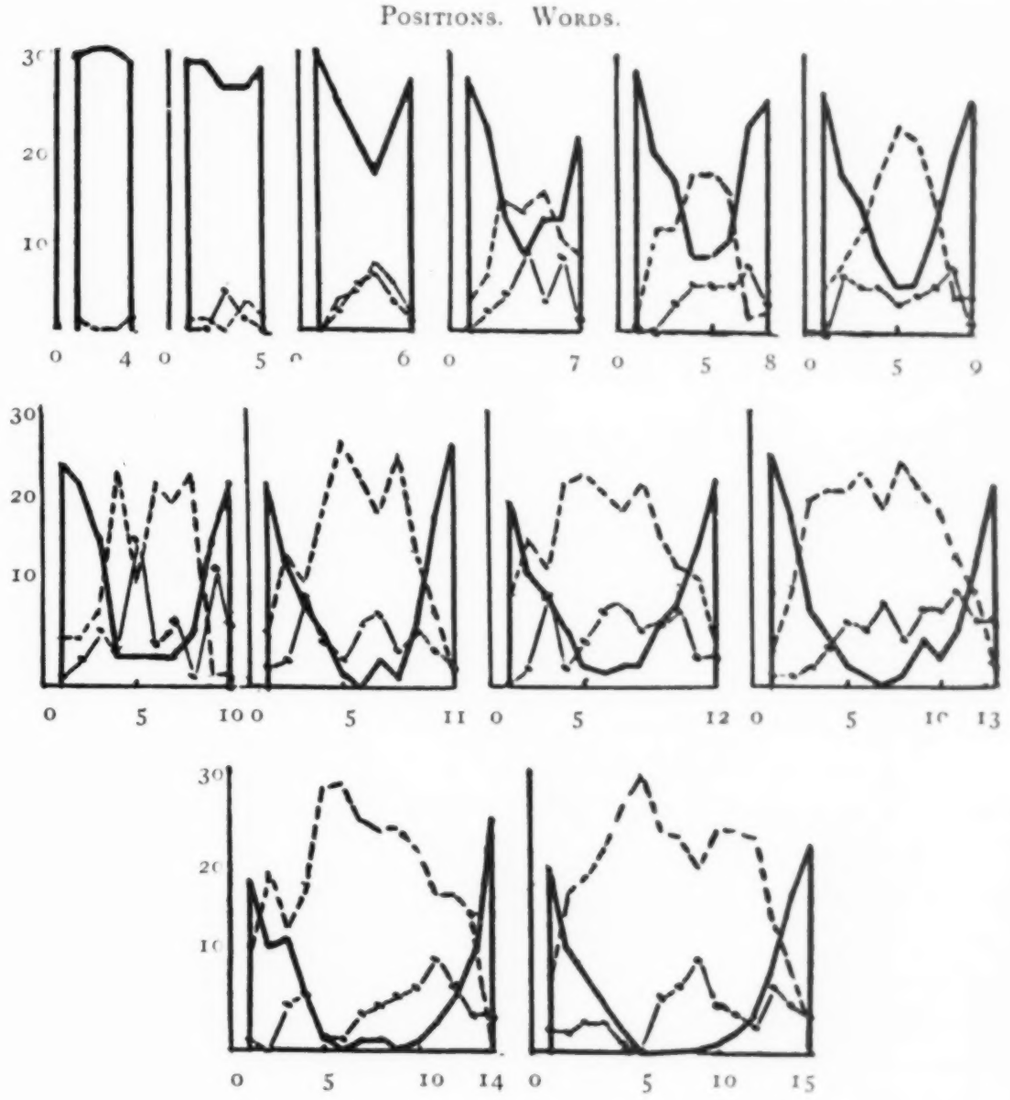

Fig. 6. Curves showing effect of quantity on distribution of positions correctly reproduced.

Continuous Lines = positions correctly placed.
Dots and dashes = positions misplaced.
Dashes only = positions omitted.
Ordinates = frequencies of reproductions.
Abscissas = order of positions in each series.

less than the memory-span is a zero abscissa line, and, as the quantity increases, this line forms a little lump in the middle

which grows larger and larger as the stimulus is increased, until the high point embraces all the positions in its ordinate; and from this, the line gradually falls toward the base at each end. Third, the misplaced positions: The distribution of these is irregular, but their tendency is to follow omissions.

The essential significance of these conclusions on the effect of quantity upon immediate reproduction is that the mind in learning a quantity of material must proceed by steps just as the body does in covering a quantity of space. The extent of that step is fixed just as much as the bodily step, and just as the body falls flat when it tries to take too large a step, so the mind literally falls when it tries, in one act, to comprehend a material that is much beyond its memory-span. It simply does not get it, and, in fact, it gets less than if it proceeded by small steps.

The significance of this for the economy of learning is evident. It means that the mind can learn only by parts, no matter what the method of reading is. If the whole method is followed, it means that different parts come above the memorial threshold at different times. If the part method is followed, fewer readings will be required to bring the part above this threshold. However, a correct reproduction must reproduce not only the members but also the order of the series. We have seen that learning in quantities greater than the memory-span reduces the learner's capacity for this kind of reproduction. Is it not then a saving of energy and time to adjust the optimum learning unit to the learner's memory-span? It appears that the above facts are a fatal blow to the much puffed-up whole method as such, and that the part method is much more in agreement with psychological laws. If so, the economy of learning as regards method is how to avoid conflicting associations.

THE INFLUENCE OF COLOR ON APPARENT WEIGHT. A PRELIMINARY STUDY

BY J. E. DE CAMP

Stanford University

CONTENTS

	Page
I. Statement of the Problem	347
II. Literature	348
III. Experimental Work:	
Section I	349
A. Apparatus	350
B. Method: Part 1	351
Method: Part 2	353
C. Data and Results	355
Section II	360
IV. Discussion	367
V. Conclusions	370

I. STATEMENT OF THE PROBLEM

The apparent weight of an object is a function of its size. Two objects of like material, appearance, etc., and having the same weight, but of different size, will appear unequal in weight not only to the unsuspecting subject, but as well, though to a less marked extent, to the experienced individual who is thoroughly familiar with the nature of the illusion involved. A second psychological factor influencing the weight of an object has been pointed out by Seashore.[1] This factor is the material out of which the object is thought to be constructed. The apparent weight of an object is influenced by many other factors than its size and material. It depends upon such factors as the neuro-muscular adjustment brought about by lifting one or more weights, the time elapsing between this adjustment and the lifting of the weight to be judged, the velocity with which the weight is raised from its support, the time occupied in the lifting process, whether it

[1] Univ. of Iowa Studies in Psychol., 1899, 2, 36–46.

is lifted before or after the weight with which it is to be compared, bodily conditions such as fatigue, etc.[1]

The question arose in the mind of the writer whether the color of an object might not influence its apparent weight. Would objects alike in all respects save that of color appear equal in weight? Is color in any way a determinant of the apparent[2] weight of an object? If the color of an object does influence its apparent weight we should like to know how the latter varies with a variation in hue, tint, and chroma. This paper and the experimental work herein reported will serve only as an approach to an answer to these questions. Its main object is to suggest, with experimental evidence for the suggestion, an answer to the particular question: Is color in any way a determinant of the apparent weight of an object? In other words, is there a color-weight illusion just as there is a size-weight illusion and a material-weight illusion?

II. Literature

So far as the writer is aware no experimenter has given a careful consideration to the problem with which we are now concerned. Seashore,[3] in connection with his experiments on the material-weight illusion, made a few tests on children to determine the presence or absence of a color-weight illusion. His method was that of selecting from a graded series a block thought to be equal to a comparison block. Only black and white were tested. His results indicated the absence of any color-weight illusion. It is obvious that the increment used

[1] Compare in this connection C. E. Seashore, 'Psychology in Daily Life,' N. Y., 1913, 175 ff.; G. E. Müller and F. Schumann, *Arch. f. d. gesamm. Physiol.*, 1899, 45, 37–112; L. J. Martin and G. E. Müller, 'Zur Analyse der Unterschiedsempfindlichkeit,' Leipsig, 1899.

[2] *Psychological* weight might be substituted for *apparent* weight. The qualification is necessary to distinguish the apparent weight when lifted by an experiencing person from the physical weight of the object. I have often noticed that the beginner in psychology not infrequently has difficulty in immediately realizing that the psychological weight of an object is just as real as its physical weight. The fact that reference is always made to a physical standard leads to the use of the term illusion. The pound of lead is really psychologically heavier than the pound of feathers. Scripture's law or formula (*Science*, N. S., 1897, 5, 227), of the size-weight illusion is only a means of passing from one real to another.

[3] Univ. of Iowa Studies in Psychol., 1899, 2, 45.

(5 grams) in his standard series of blocks was too large to separate out any factor whose influence is very slight. A second approach is some work done by E. Bullough[1] on 'The Apparent Heaviness of Colors,' which is, as the writer of the article points out, a contribution to the æsthetics of color. He is concerned in particular with the æsthetic effect of placing one color above another. It may be said that he attempts to investigate scientifically the rule "that dark colors should stand below light ones."[2] His method is that of choices or preferences. It is concluded that dark colors are heavier than light ones. According to him "it is the luminosity which forms the principal factor in the preferences,"[3] *i. e.*, the relative luminosity of a color determines whether it will be preferred above or below another color. Seemingly the empathic sense in which the term weight is here used differs quite markedly from that in which we use the term weight in speaking of a lifted object, but possibly the significance of the term in the two cases is quite the same. Our ignorance of the factors involved in the æsthetic arrangement of two or more colors at present precludes a definite answer to this question.

There are various articles dealing with the many factors influencing the apparent weight of a lifted object, such as the size of the object, the rate at which it is lifted, etc., but since we have endeavored to make such factors constants in our tests they need not claim our attention here, and it is not necessary to refer to the literature dealing with them.

III. Experimental Work

Section I

The experimental work reported in this section was carried on by the writer and two of his advanced students[4] in

[1] *British J. of Psychol.*, 1907, 2, 111–152.

[2] *Loc. cit.*, 113.

[3] *Loc. cit.*, 133; *cf.* also p. 152 where he says: "Since the saturated colors are not all of the same luminosity, but yellow and green are of higher luminosity than blue and violet, yellow and green are also of apparently lighter weight than blue or violet when seen singly."

[4] Misses Irene Hollenbeck and Laura Stephan. Sixty-five per cent. of the tests were made by the writer.

the psychological laboratory of the University of California during the second semester of the year 1915–1916. Of the sixty-three subjects tested nineteen were men and forty-four were women. The subjects were drawn from Professor Stratton's class in applied psychology. Seven of the subjects were seniors, thirty-five were juniors, nineteen were sophomores, and two were specials. The test was always given in the forenoon some time between 8 A. M. and 12 M. Due to a cloudy or foggy morning there were a few instances where the light conditions were slightly below normal (ordinary diffuse daylight). The test usually occupied about forty-five minutes, and was given to each subject individually. Each test fell into two main parts: for convenience we will call the first part Part 1, and the second part Part 2. Part 1 always preceded Part 2.

A. Apparatus.—The apparatus used was twenty-five cubical ($2\frac{1}{2} \times 2\frac{1}{2} \times 2\frac{1}{2}$ in.) soft pine blocks. These blocks were smoothed and made the desired weight by boring out part of the wood from the bottom of the block. Nine of the twenty-five were used in Part 1, all were used in Part 2. The nine blocks for Part 1 were all made exactly the same weight, seventy-six grams. Each one of the nine was covered with colored paper by glueing the paper evenly and closely to five sides of the block. The bottom was left uncovered. In this way the nine blocks were alike in size and weight; in fact they were alike in every respect save that of color. Each of the nine differed from every other one in respect to color. The following colors were represented: red, orange, yellow, green, blue, violet, purple,[1] black, and white. It will be seen later that the present work does not demand a refined photometric method such as is described by Sir Wm. Abney,[2] or like that given by Yerkes and Watson.[3] An attempt was made to obtain the approximate luminosity value of each color by the method suggested by Rood.[4] The following values were received for the luminosity of the different

[1] Probably more correctly designated by the commercial term *cerise*.

[2] 'Researches in Color Vision,' London, 1908.

[3] 'Behavior Monographs,' 1911, No. 2.

[4] *Amer. J. of Sci. and Arts*, 1878, **15**, February.

colors: red, 5.2; orange, 27.2; yellow, 75.7; green, 44.3; blue, 14.7; violet, 8.5; purple, 6.6; black, 0; white, 100; gray, 16.9. The figures given represent the percentage of white contained in the colored discs. No pretense is made to absolute accuracy. There can be little doubt, however, that the relative luminosity of the different colors is correctly shown.

The sixteen additional blocks used in Part 2 were of the same material and size as those used in Part 1. These were covered on five sides with gray paper (No. 16[1] of Hering's series). They all differed in weight. The sixteen made up a graded series from fifty-five to one hundred grams, the increment in each case being three grams. The set of sixteen blocks were alike, then, in every respect save that of weight.

Additional apparatus consisted of (*a*) a stand supporting a horizontal rod 16 cm. above the base of the stand, (*b*) a second stand supporting a black cardboard screen, and (*c*) a large piece of soft black velvet cloth. The horizontal rod mentioned under (*a*) was wrapped with black cloth.

B. Method: Part 1.—The general method used in Part 1 was that of Paired Comparisons. This method is so well known that it needs no discussion here.[2] The ranks and files of the table constructed for the thirty-six (*i. e.*, between nine objects; $\frac{9 \times 8}{2} = 36$) possible comparisons were headed thus:

	R	O	Y	G	Bl	V	P	Bk
O	1							
Y	2	3						
G		4	etc.					
Bl								
V								
P								
Bk								
W								

The observer[3] was seated at one side of a fairly large table, the experimenter at the other. The table was covered with the large piece of black velvet cloth. The stand supporting

[1] W. Brown has shown (PSYCHOL. REV., 1915, **22**, 520 f.) that this particular gray is the seventeenth (instead of the sixteenth as numbered) in Hering's set.

[2] See E. B. Titchener, 'Experimental Psychology,' Vol. 1, Pt. 1, 92 f.

[3] Throughout this paper the terms *subject* and *observer* are used interchangeably.

the horizontal rod was placed near the observer. The purpose of the rod was to give uniformity to the height of each lift made by the observer. The height of the rod above the table was 16 cm. The observer was definitely shown how the lift was to be made. The elbow was to be used as a fulcrum; no other part of the arm or hand was to be bent. The subject[1] was told to have as nearly as possible a uniform method of grasping the blocks. The blocks were placed in position by the experimenter for the subject. Only two blocks, those being compared, were allowed on the table at the same time; the others were kept in the table drawers on the experimenter's side of the table. The cardboard screen was placed in such a way that the protocol of the experimenter was concealed from the subject's view.

The first comparison was made between red and orange, the second between red and yellow, the third between orange and yellow, etc. Care was taken that in the eight comparisons belonging to any color the block representing the color in question should be presented four times before and four times after the blocks with which it was compared. This served to eliminate any temporal error. All blocks were lifted from the same position, in that way avoiding any positional error. Before the test began the observer was impressed with the following four points:

1. The experiment is a comparison of weights.

2. The weights are all different; no two have the same weight.[2]

[1] *Vide* note 3, p. 351.

[2] It might be thought that the proper procedure would be to omit any reference to the weight of the blocks. In such a case should the observer be presented in the first few trials of the experiment with blocks whose color has little relative influence upon their weight he would very likely get the impression that the blocks were all equal in weight. The influence of this idea would invalidate the results obtained so far as the factor under investigation is concerned. Suppose that the different colors have little relative influence upon the apparent weight of objects, the introduction of this common factor will greatly aid in the determination of the little and will not materially affect its validity. The suggestion will aid in fixing in the mind of the observer the idea that the aim of the experiment is a comparison of weights and will tend to exclude from him any knowledge of the factor under consideration. This is highly desirable since the untrained observer can not avoid distorting actual occurrences; his results are biassed by what he thinks the results ought to be under such conditions.

3. The second weight in each comparison is to be judged in terms of the first.

4. The eyes are to be kept fixed upon the block while it is being lifted.

After these instructions had been given a few preliminary trials were usually made with some of the blocks with the graded gray series for the purpose of (1) familiarizing the observer with the experimenter's method, and (2) setting up a kind of muscular adjustment in each individual case for the perception of differences in weight. At first fairly large differences were given, then smaller ones. In these preliminary tests the subject was informed whether his judgments were right or wrong.

Part 1 was then conducted in accordance with the method of paired comparisons. The observer was in no case informed whether his judgment in any particular instance was correct. Sometimes when the observer showed too much anxiety about the correctness of his judgments the experimenter made such casual remarks as, "You are doing very well," or, "You are getting along very nicely." Immediately upon each individual judgment the experimenter entered the judgment in the table previously prepared for that purpose. Unless the observer showed signs of fatigue the test was carried directly through. This part of the test required approximately fifteen minutes. In case of fatigue, and this very seldom happened, the observer was allowed a brief respite.

Method: Part 2.—In this part the complete set of twenty-five blocks was used. Only two blocks were on the table at a time. Beginning with the red block and taking the colored blocks in spectral order each one served as a standard with which a number of the sixteen gray blocks were compared. In each case the colored block was lifted first, then the gray block. The gray blocks were taken in the order of their weight, *i. e.*, 55, 58, 61, . . ., 100. Before beginning this test the observer was told that in this case some of the blocks compared might be just alike in weight, in this way indicating to the observer that in some of the comparisons 'the same' might occur as a correct judgment as well as 'lighter' or

'heavier' in other comparisons. The observer was not informed that the gray blocks constituted a graded series.[1] Record was made by the experimenter of the subject's judgments. In the first few experiments the comparison of any colored block with the graded series of grays was stopped upon the subject's first judgment of 'heavier.' In the later tests the comparison was stopped only when one judgment of 'heavier' was followed in the next comparison by a second judgment of 'heavier.' It was assumed that a continuation of the series would lead to a judgment of the remaining heavier weights as 'heavier.'

After the completion of Part 2 four of the colored blocks, red, yellow, green, and blue, were placed before the observer ranging from his left to right in the order named. He was told to lift these blocks separately as many times as he desired, no horizontal rod being used to restrict the height of each lift, and to place them in the order of their weight, the heaviest one to his left, the next heaviest one place to the right, and so on. By means of a stop-watch was recorded the time that it took the observer to arrange these four blocks in the manner described.

All nine of the colored blocks were then placed in chance order before the subject. He was told to select that block which was most pleasing, *i. e.*, whose color was most pleasing, also to select the most pleasing block of the remaining eight, the most and the next most displeasing blocks.

In no case, not even at the close of the test, was the subject told the purpose of the experiment or the point under investigation. No information was given the subject about the results obtained in the experiment.

[1] This varies slightly from the method recommended by Whipple, 'Manual of Mental and Physical Tests,' Baltimore, 1915, Pt. 2, 226, for a similar effect in the case of the size-weight illusion. In his case the graduation while not distinctly mentioned is open to easy discovery. In Part 2 of the present experiment, following upon Part 1, in which no apparent (to the observer) order is followed, no impression is conveyed that the grays are given in any order other than a promiscuous one. Doubtless the repeated giving of the graded series with the colored blocks led to an idea, on the part of the observer, of the gradation, yet in no case, to the knowledge of the writer, was any indication given that the gradation was noticed. With the results of this part of the experiment at hand it seems probable that more reliable results would have been obtained had the subject been informed that the series was a graded one.

The subject's name, sex, year in the university, and the condition of the light were indicated upon the record sheet.

C. Data and Results.—The results obtained by the method of paired comparisons (Part 1) are graphically shown in Fig. 1.

Arranging the colors in the order of their apparent heaviness we have red, white, orange, violet, green, purple, blue, yellow, black. In this decreasing series three characteristic step effects may be clearly seen. The highest is formed by the two colors red and white; the next step, whose height corresponds fairly closely to the line of chance, may be thought of as made up of orange, violet, green, and purple; the third step is made up of blue, yellow, and black. Stated in another way, the results seem to indicate that red and white have a positive; blue, yellow, and black a negative; orange, violet, green, and purple an indifferent; effect upon the apparent weight of the objects lifted.

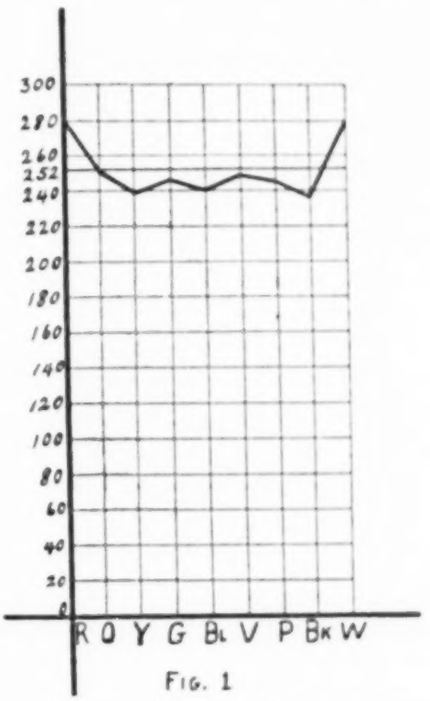

FIG. 1

FIG. 1. Method of paired comparisons. The results for 63 subjects are shown. The ordinates represent the total number of judgments of 'heavier' given the different colored blocks.[1] The different colors are represented along the line of abscissas. The height to which a color should rise, assuming the absence of all factors save that of chance, is shown by the horizontal line at ordinate 252.

The average value,[2] the average value in terms of per cent.,

[1] Though no encouragement was given to the subject to reply that the blocks compared were of the same weight such judgments occasionally occurred. In cases of this kind, also in those cases where the subject seemed quite doubtful, the comparison was repeated, but in the curve here shown these cases were disposed of by giving to each color involved one half of one unit credit, it being the plan to give to the color judged 'heavier' one unit credit.

[2] The 'average value' for each color is obtained by dividing the total number of times for all subjects it was judged 'heavier' by the number of subjects. If any color had always received a judgment of 'heavier' regardless of the color with which it was compared its average value would be 8. Assuming no influence other than chance the average value for any color would be 4. The M. V. as given in the table is computed from the average.

and the mean variation for each color are given in the following table, Table I.:

TABLE I

Color	Av.	%	M. V.
Red	4.452	55.6	1.083
Orange	4.008	50.0	0.899
Yellow	3.794	47.4	1.203
Green	3.905	48.8	1.011
Blue	3.810	47.6	1.180
Violet	3.960	49.5	0.943
Purple	3.897	48.7	1.194
Black	3.754	46.9	1.218
White	4.421	55.3	1.194

Table II. is presented to show the gross distribution about the chance line of the individual results obtained. The results, for the men, for the women, and for the men and women are shown in separate columns.

TABLE II

Color	Men			Women			Men and Women		
	>C	<C	=C	>C	<C	=C	>C	<C	=C
Red	12	4	3	22	13	9	34	17	12
Orange	5	9	5	17	15	12	22	24	17
Yellow	4	10	5	15	22	7	19	32	22
Green	8	6	5	13	22	9	21	28	14
Blue	7	8	4	12	20	12	19	28	16
Violet	7	9	3	16	11	17	23	20	20
Purple	8	7	4	14	21	9	22	28	13
Black	5	8	6	14	25	5	19	33	11
White	10	5	4	23	16	5	33	21	9

In the columns headed > C is given the number of individuals whose credit value for the particular color was greater than chance would have allowed. In column < C (= C) is given the number of individuals whose credit value for the different colors was less than (equal to) that allowed by chance.

Considering the totals we note the same step-effect as was pointed out above, and each step is made up of exactly the same colors as before: red and white make up the first; orange, green, violet, and purple, the second; yellow, blue, and black, the third. Even though the number of men tested is quite too small to merit serious consideration of

their results apart from the combined results of both men and women it will be seen that save for orange and blue their distribution is quite in accordance with the statements made above.

If we now note the results obtained from the auxiliary test[1] with the four colors, red, yellow, green, and blue, we have the following table, Table III.:

TABLE III

	Red		Yellow		Green		Blue	
First place	25	13	11	13	10	13	6	13
Second place	16	13	11	13	12	13	13	13
Third place	7	13	17	13	19	13	9	13
Fourth place	4	13	13	13	11	13	24	13

First place (second place, etc.) is used to denote the fact that a block falling therein was designated by the observer as heaviest (next heaviest, etc.) of the four. The first column under each of the four colors indicates the number of individuals giving to the various colors the first, second, third, and fourth place. The second column gives the distribution according to chance.

If we adopt an arbitrary system of units in which the occurrence of a colored block in the first place is given a value of 4; its occurrence in the second place, a value of 3; in the third place, a value of 2; in the fourth place, a value of 1; and compute upon this basis the aggregate value for each of the four colors, we have: red, 166; yellow, 124; green, 125; blue, 105. The figure of chance is 130. These results, save those for yellow, quite accord with our previous statement (see p. 355). To yellow is attached a slightly larger figure than we should have anticipated on the basis of our previous results. To agree with the previous results it should have attained a figure near that of blue, about 105. This slight discrepancy may be due to the presence in this part of the test of complicating factors arising out of the increased freedom in method allowed the subject.

The average time spent in arranging the four blocks in

[1] The method of this test is given on page 354 *supra*.

the order of their apparent weight was 1 min. 59 secs. This result is obtained by averaging the results of forty-seven subjects. The experimenters began to take the time with the sixteenth subject in the series, and failed to get the time in a single subsequent case.

The test requiring a choice of the two most pleasant and the two most unpleasant of the nine blocks led to the following results, Table IV.:

TABLE IV

	Pleasant		Unpleasant	
Red	2	4	1	4
Orange	4	8	6	9
Yellow	4	4	3	5
Green	7	14	3	4
Blue	14	11	1	4
Violet	7	2	4	9
Purple	6	5	20	11
Black	0	1	12	2
White	6	1	0	2

In the first column under Pleasant (Unpleasant) appears the number of persons choosing red (orange, yellow, etc.) as the most pleasant (unpleasant) color of the nine. In the second column under Pleasant (Unpleasant) is given the number of persons choosing red (orange, yellow, etc.) as the next most pleasant (unpleasant) color. From the table it readily appears that the most generally pleasant color was blue, with green next; the most generally unpleasant color was purple, with black next. Three of these colors, black being the one excepted, showed no particular influence upon the block so colored. Red and white called out little affective quality, and it was these two colors[1] that seemingly had the most marked effect upon apparent weight. Again, of the three colors, yellow, blue, black, which appeared to influence apparent weight negatively, one (blue) is generally pleasing, one (black) is rather displeasing, the other (yellow) is neither decidedly pleasant nor decidedly unpleasant.

The following table, Table V., shows the results obtained from the second part of the experiment (designated above, Part 2):

[1] *Supra*, p. 355.

TABLE V

Red	68.27	Violet	70.78
Orange	68.32	Purple	70.07
Yellow	67.46	Black	69.57
Green	68.22	White	70.56
Blue	68.30		

The figures opposite the various colors are obtained in the following manner: for each individual test an interpolated value for each color is computed by averaging the value of the last judgment of 'lighter' and that of the first judgment of 'heavier'; the sum of these individual interpolated values is then averaged, and it is the resulting figure that appears in the table. Probably little of a definite nature is to be gained from a consideration of this table. It seems that the method was too crude for satisfactorily disengaging the factor under consideration from the many others which complicate the problem of the apparent weight of a lifted object. Doubtless one of the main complicating factors was the fact that the colored blocks towards the end of the series would not be approached with the same naïveté as were those of the first part of the series. It will be noted that there is a decided tendency to overestimate the weight of the gray block. This overestimation is due, I take it, to the serial arrangement of the grays.[1] It is the result, often noted in the case of unsophisticated observers, and clinging even to the trained observer, of the premature equalization, when there is a problem of equalization, of members of a series of diverse stimuli with a standard stimulus. Seemingly in all such cases there is a tendency to judge that the goal is reached *before* it is reached rather than to judge that the goal is reached *after* it has been reached. This tendency is quite noticeable in the various experiments connected with the problem of the limen. The following is sufficient evidence that this factor was really an effective one in the present experiment. Though each gray block was always heavier than the preceding one very frequently the observer would

[1] Subsequent experiments (Section II.) clearly show that this overestimation is not due to the color (gray) of the block. It may be that the grayness has some influence, but it is relatively small. Compare results given on p. 363 *infra*.

judge a very light block (say one of 64 grams) to be equal to or heavier than the colored block (weight, 76 grams), then judgments of 'lighter,' 'the same,' and 'heavier' would follow. Only one of the sixty-three individuals tested failed to definitely show this tendency, and it was not an uncommon occurrence for this tendency to appear more than once during any particular test. In the opinion of the writer anticipatory attention has entered as a complicating factor in the averages presented. If so, its influence should be stronger for those colors appearing first in the series, since numerous comparisons would in all probability tend to lessen or altogether eliminate it in the case of those colors coming towards the end of the series. Were this factor alone present it seems that the series of values from red to white should show a gradual increase. It is perfectly clear that in the figures this increasing tendency is present. A separate average for the first, second, and third group of three values gives 68.02, 69.10, and 70.07 respectively. But it is quite probable that anticipatory attention is not the only influence present, and the results of our test by the method of paired comparisons, which seems to be the most adequate method of attacking the problem in hand, would lead us to assume an influence of the different colors. The low value (lowest of the nine) of yellow accords with the results previously obtained.

Section II

This section contains a report of additional experiments carried out with the hope of obtaining more satisfactory results relative to the existence of a color-weight illusion. The work was done at Stanford University during the year 1916–1917. The subjects used, sixty-eight men and thirty-one women, were drawn from various classes in psychology, but mostly from my class in mental hygiene. First-, second-, third-, and fourth-year students were represented.

While the apparatus and method employed were similar to those used in the first part of Section I., a brief description is necessary for a clear understanding of the results to follow.

Apparatus.—Six cubes ($2\frac{1}{2} \times 2\frac{1}{2} \times 2\frac{1}{2}$ in.) were constructed

from ordinary Bristol board; the bottom of each was left open; the remaining five sides were covered over with a colored paper, the color varying with the cube. The six colors represented were red, yellow, blue, black, white, and gray. These colors were the same as those used in Section I. with the exception of the gray, which in the present case was No. 17 of Hering's series. (No. 16, the gray previously used, was not available.) Each cube was made to weigh seventeen grams. Adjustment of weight was accomplished by glueing a cork to the bottom of the upper side of the cube and adding shot or paring away cork. The choice of a small absolute weight for the cubes was dependent upon the consideration that if the color of an object has an influence upon its apparent weight that influence should be relatively greater the lighter the object, since the lighter the object the smaller the appreciable difference in weight.

Method.—The subject was seated at one side of a table, the experimenter at the other. The following experiments were carried out:

1. The four cubes, red, yellow, blue, and white, were placed on a large sheet of dark cardboard lying upon the table. The right-to-left arrangement of the cubes varied from subject to subject. The following four orders of presentation were used, beginning at the subject's right: (1) red, yellow, blue, white, (2) yellow, blue, white, red, (3) blue, white, red, yellow, and (4) white, red, yellow, blue. The first (1) represents the order in which the cubes were placed before the first subject; for the next subject the order was changed to that shown in (2); etc.

The subject was told that the cubes were all different in weight and that he should arrange them in the order of their weight, placing the heaviest at his left. He was definitely instructed to look at the weights while lifting them. An occasional prompting was sometimes found necessary to make the subject conform to this requirement. The subject was allowed to rearrange the order[1] as often as desired until

[1] A change in the order often occurred. One of the most frequent means of bringing about such a change was the following: the subject would lift the various

the proper order was obtained. The subject's final arrangement was noted and recorded.[1]

2. A test similar to the one just described was given using the three cubes, white, gray, and black.

3. The subject was tested with cubes red, yellow, blue, white, and black, using the method of paired comparisons. This method has already been referred to under Section I. (p. 351). The same general precautions were taken here as there.

4. In some of the later tests the subject was asked to designate from 'looks' alone the order of the apparent heaviness of the cubes. After the subject had given his order he was asked to explain why the cube so designated looked heaviest and another seemed lightest.

The results obtained for Test 1 appear in Tables VI. (*A*) and VI. (*B*).

TABLE VI. (*A*)

Color	Men (68)				Women (31)				Men and Women (99)			
	1	2	3	4	1	2	3	4	1	2	3	4
Red	18	16	19	15	6	5	13	7	24	21	32	22
Yellow	15	19	22	12	8	10	8	5	23	29	30	17
Blue	17	20	17	14	9	10	5	7	26	30	22	21
White	15	13	10	30	8	6	5	12	23	19	15	42

Under 1, 2, 3, and 4 appears the number of times the different colors were ranked as heaviest, second heaviest, third heaviest, and lightest. The numbers in parentheses following Men, Women, and Men and Women indicate the number of subjects.

Assigning the following values to the different judgments, heaviest, 4; second heaviest, 3; third heaviest, 2; lightest, 1; and computing the aggregate value for each color we have Table VI. (*B*).

The significance of these results is not very clear. They

cubes one or more times and place them in a certain order, then he would compare the extremes, seemingly as a confirmatory test, and that would frequently lead to a change in the established order.

[1] This method will be referred to later as the 'arrangement' method.

TABLE VI. (*B*)

Color	Men (68)	Women (31)	Men and Women (99)
Red	173	72	245
Yellow	173	83	256
Blue	176	83	259
White	149	72	221

do not support the interpretation given to the results of Section I., neither do they accord with the results obtained by the method of paired comparisons referred to later in this section (see p. 364). There seems to be a tendency to judge white and red as relatively light, blue is heaviest, while yellow follows as next heaviest. It is to be noted that there is not complete agreement with the rule: light colors—light weight, dark colors—heavy weight.

The results for Test 2 are given in Tables VII. (*A*) and VII. (*B*).

TABLE VII. (*A*)

Color	Men (68)			Women (31)			Men and Women (99)		
	1	2	3	1	2	3	1	2	3
White	23	16	29	10	5	16	33	21	45
Gray	23	30	15	9	12	20	32	42	35
Black	23	22	23	12	14	15	35	36	38

Under 1, 2, 3 appears the number of times the different colors (grays) were ranked as heaviest, next heaviest, and lightest.

Assigning the following values to the judgments, heaviest, 3; next heaviest, 2; lightest, 1; and computing the aggregate value for each color we have Table VII. (*B*).

TABLE VII. (*B*)

Color	Men (68)	Women (31)	Men and Women (99)
White	186	56	242
Gray	205	69	274
Black	215	71	286

From these results for three blocks differing only in tint it is quite apparent that the tendency is to judge the darker block as heavier. This is true for both men and women.

The results for Test 3 appear in Table VIII.

TABLE VIII

Color	Men (42)	Women (8)	Men and Women (50)
Red	97	23	120
Yellow	90	21	111
Blue	96.5	29	125.5
Black	96.5	30	126.5
White	101	23.5	124.5
Gray	101	25	126

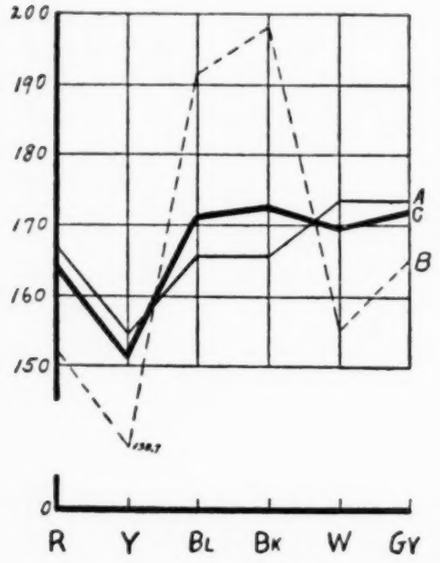

FIG. 2. Curve *A* indicates the relative difference in the apparent weight of six different colored cubes of the same weight, as judged by 42 men. Curve *B*, the same for 8 women. Curve *C* is obtained by combining the results for both men and women. The colors are represented along the line of abscissas, the relative number of judgments of 'heavier' along the line of ordinates.[1] Method of paired comparisons.

The numbers represent the relative number of judgments of 'heavier' given in a comparison of different colored cubes. A graphic representation of these results is shown in Fig. 2.

For the men white, gray, and red are the heavier colors, yellow is the lightest; for the women[2] black, blue, and gray are the heavier, as with the men yellow is the lightest. White is heaviest for the men while for the women it is slightly heavier than black and red.

Since the method of paired comparisons was used in the experiments of both Sections I. and II. it was deemed advisable, for the sake of increasing the number of comparable results from like tests, to combine the values received in the two. The colors common in the two groups were red, yellow, blue, white, and black. The judgments resulting from a

[1] In this figure, as well as in all subsequent ones, the absolute ordinate values appertaining to any individual curve have been transformed so that their summation equals 1,000.

[2] The number of women tested is too small to make sure that their results are typical.

comparison of all possible combinations of these five colors were separated out from the two wider groups of results belonging to Sections I. and II. Table IX. presents the results obtained.

TABLE IX

Color	Section I			Section II			Sections I and II		
	Men	Women	Men and W.	Men	Women	Men and W.	Men	Women	Men and W.
	(19)	(44)	(63)	(34)	(8)	(42)	(53)	(52)	(105)
Red	44.5	109	153.5	69	16	85	113.5	125	238.5
Yellow	34	82	116	67	15	82	101	97	198
Blue	32	78.5	110.5	65.5	17	82.5	97.5	95.5	193
Black	39.5	88.5	128	70	18.5	88.5	109.5	107	216.5
White	40	82	122	68.5	13.5	82	108.5	95.5	204

The table brings together the results, method of paired comparisons, of Section I., Section II., and Sections I. and II. for red, yellow, blue, black, and white. In the different columns appear the values given by men, women, and men

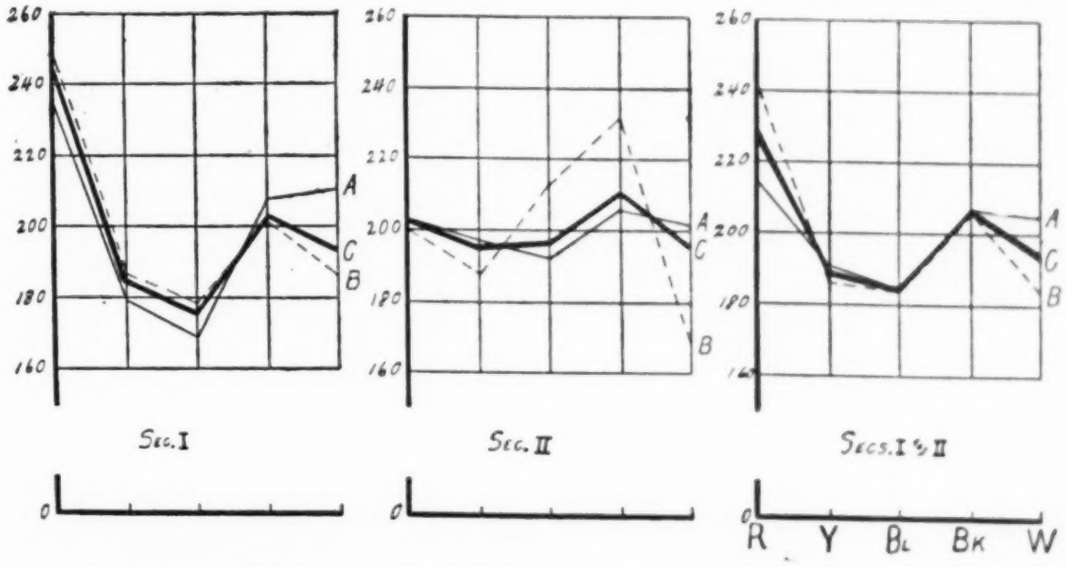

FIG. 3. Method of paired comparisons. The results of Sections I., II., I. and II. combined, are shown graphically. In each case Curve *A* represents the apparent relative weight of the five colored cubes, red, yellow, blue, black, and white as judged by the men. Curve *B*, the same for the women. Curve *C* represents the combined results for men and women. The colors are shown along the line of abscissas; the relative number of judgments of 'heavier' along the line of ordinates.

and women for each color. The numbers in parentheses designate the number of subjects.

For the purpose of facilitating comparison the results of Table IX. are graphically represented in Fig. 3.

It will be seen that there is great similarity in the general direction of the various curves. The results of both Sections I. and II. seem to indicate that there is a general tendency to judge red and black as relatively heavy, white fairly heavy, and yellow and blue light. The main departure from this tendency occurs in the women's judgment of white. For them white occupies a low position similar to that taken by yellow and blue, red and black being the heavy colors. Attention may be called to the fact that for the women red is noticeably the heaviest of all.

The results for Test 4 appear in Table X.

TABLE X

Color	Men (12)	Women (4)	Men and Women (16)
Red	44	15	59
Yellow	23	9	32
Blue	50	16	66
Black	56	23	79
White	25	8	33
Gray	44	13	57

The numbers represent the relative influence of 'looks,' without actual lifting, upon the different colored cubes. These numbers are obtained by assigning the following values to the different judgments, the block judged heaviest, 6; next heaviest, 5; . . .; lightest, 1; and computing the total credit due each color.

A graphic representation of the results appearing in Table X. is shown in Fig. 4.

The results for Test 4 are much more uniform than those received in the other tests. The uniformity from subject to subject rendered the testing of a larger number of subjects unnecessary. The relative position for each of the six colors used is the same for both men and women, save in the case of yellow and white which show a small difference. These results favor the rule: light-colored objects appear light in weight; dark-colored objects appear heavy in weight.

The introspections taken in connection with the judg-

ments of Test 4 indicated quite clearly that the judgments made were due in most cases to associations of a general and not of a specific nature. For example, the connection between the black cube and heavy weight was immediate, because dark objects are usually thought of as heavy, and did not depend upon the arousal of some specific association, like the idea of coal or iron.

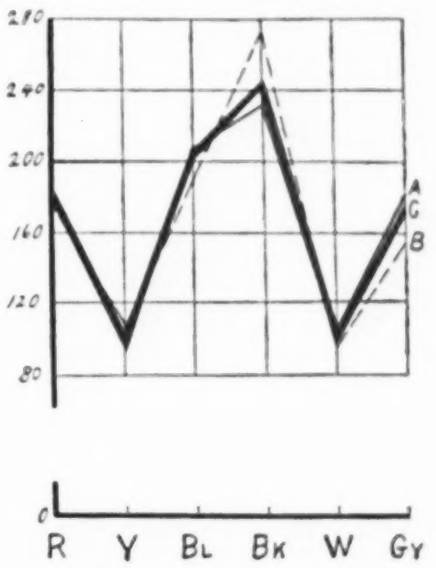

FIG. 4. Curve *A* indicates the relative difference in the apparent weight, judged by 'looks' alone, of six different-colored cubes as judged by 12 men. Curve *B*, the same for 4 women. Curve *C* represents the combined results of both men and women. The colors are shown along the line of abscissas. The ordinates represent the relative weight of the different colors.

We raised an objection to accepting at their face value the results of Part 2 of Section I. because of the probable influence of anticipatory attention. The values for the colors, red, yellow, blue, black, and white (the colors for which we have extended results with the method of paired comparisons), do not vary widely in their distribution from the distribution obtained by the method of paired comparisons. If the value for red were slightly greater (and if we allow for anticipatory attention it would be relatively larger) the curve of distribution for these values would be quite similar to those obtained by the method of paired comparisons. The low position of yellow is to be noted. If anticipatory attention is not admitted the results for blue in the two cases do not accord.

IV. DISCUSSION

The results accruing from the problem herein attacked probably need little further discussion. They indicate that when the method of paired comparisons is used there is a general tendency to judge the apparent weight of red and black as heavier than that of yellow and blue. The fact

that red and black always lie above,[1] yellow and blue below, the line of chance (at ordinate 200) may be interpreted as due to the overestimation of the weight (judged after lifting) of the colors red and black and the underestimation of yellow and blue.

The results received when the arrangement method was used (p. 362 f.) are not in harmony[2] with the above interpretation. They seem to indicate quite clearly that white is judged lighter than red, yellow, or blue.

There is no doubt in the writer's mind that the results obtained by the method of paired comparisons must be considered as more reliable than those obtained by the arrangement method. The former appears to be the more scientific method. A number of distinct judgments is involved and the situation immediately preceding each judgment is comparatively simple. When the arrangement method is used the placing of all the blocks before the subject renders the situation preceding his judgments rather complex. That the method of arrangement permits of rearrangements is probably an argument against it. Often the cause of the rearrangement seems to be extrinsic and not dependent upon the appearance of the cubes (see note, p. 361 f.).

The fact that the results of the two methods, that of paired comparisons and that of arrangement, the cubes being actually lifted, show no definite correlation[3] with the results received when the judgments were made by 'looks' alone seems to indicate that the lifting of the cubes may have introduced an additional factor. One might be disposed to draw the inference that this factor is not entirely dependent upon the appearance of the cubes.

If the claim that the results obtained by the method of paired comparisons demands prior consideration is admitted, the question arises, To what is the differential influence indicated due? The fact that the effect produced is quite analogous to that present in the size-weight illusion suggests the

[1] In Curve *B* of Fig. 3 (Section II.) the ordinate for red is exactly 200.

[2] That is, if they are taken as at all indicative.

[3] With the exception of the results for blue there is a fair degree of correspondence between the curves of Figs. 3 and 4.

possibility of a similar explanation. The size-weight illusion is usually explained on the basis of a close association between large objects and heavy weight, and between small objects and light weight. Making use of a similar argument we should expect yellow and white, since they *appear* light, to be overestimated after lifting. It may be remarked that in the case of white this occurred in the experiments of Section I. But the results for yellow, the most consistent throughout, indicate just the opposite. Further, we should expect the weight of black, blue, and red to be underestimated after lifting, but this is opposite to the general tendency as shown by the results (see curves in Fig. 3). The results, then, militate against an associative explanation so simply conceived.

Since white objects are known, other things being equal, to appear larger than black objects it might be thought that a slight difference in apparent size would exist with our blocks and consequently influence their apparent weight. If the white (black) block is judged larger (smaller) than the black (white) block, and no other factor is involved, it in consequence ought to be judged lighter (heavier) than the black (white) block. This is in accordance with the general tendency of the results for white and black, but does not hold for blue. Even with white the results of Section I. indicate that it was often judged relatively heavy. Again, it has been found that 'there is a small, but decided influence of color'[1] upon the apparent size of an object. According to Quantz[2] red, orange, yellow, and purple increase, blue and violet decrease, the apparent size of an object, green is near the borderline of influence. According to these results the red object should appear larger than the blue or violet object and consequently be judged, when lifted, lighter than they, but this was not found to be the case. A similar argument with respect to some of the other colors shows the inadequacy of apparent size as a basic explanation of the results obtained. Hence we conclude that some factor other than the apparent size of the blocks is involved.

[1] J. O. Quantz, *Amer. J. of Psychol.*, 1895, 7, 40.

[2] *Loc. cit.*, 39.

There is a possibility that the different colors exert different dynamic influence upon the observer such that red and black, say, call out but little neuro-muscular force, and blue and yellow call out more of such force. This idea is quite too conceptual to be introduced into the present discussion, which already has involved too much theorizing.

V. Conclusions

The writer is fully aware that the conclusions to be drawn from the results herein presented must necessarily be tentative. It is perfectly clear that for a final solution of the present problem much careful and accurate work must be done. At the present time I only wish to suggest the following conclusions as sufficiently justified by the results obtained:

1. The influence of the color of an object upon its apparent weight is relatively slight.

2. There is a tendency in many cases, at least when the method of paired comparisons is used, to judge a red or black object to be slightly heavier than a yellow or blue object of the same weight.

3. Apparently the influence of these colors is not due to their tint value alone; it seems necessary to consider their hue as a minor factor.

4. The results obtained do not lend themselves to a simple associative explanation.

5. Seemingly there is no simple correlation between the affective quality of a color and its influence upon apparent weight.

The writer wishes to express his indebtedness to Professor Warner Brown for a helpful criticism of an early draft of this paper.

TACTUAL ILLUSIONS OF MOVEMENT[1]

BY HAROLD E. BURTT

A. Introduction

The kinematoscopic illusion in which two similar stationary visual stimuli in quick succession in different places yield an impression of movement has long been familiar. Wertheimer[2] and Korte[3] studied the phenomenon rather exhaustively with variations of intensity, distance, interval, etc. The writer has demonstrated the possibility of such illusions in audition.[4] Benussi has found similar phenomena in touch and has published two brief accounts,[5, 6] the latter of which appeared after the present study was in progress. It seemed profitable to determine whether the tactual corresponded to the other sensory fields merely in the existence of the illusion under optimal conditions, or whether there was a correspondence in the more detailed aspects. Accordingly an investigation was made of the rôle of the intensity of the tactual stimuli, the distance between them and other variables much as they were studied by Korte in the visual phenomenon.

The principal facts found by Korte are as follows:

1. If two equal straight lines or other similar visual stimuli a few centimeters apart are presented in quick succession for equal lengths of time certain conditions of intensity, time and distance yield an impression of movement from one

[1] From the Harvard Psychological Laboratory.

[2] Wertheimer, M., 'Experimentelle Studien über das Sehen von Bewegung,' *Zeitschrift für Psychologie*, 1912, 61, pp. 161–265.

[3] Korte, A., 'Kinematoskopische Untersuchungen,' *Zeitschrift für Psychologie*, 1915, 72, pp. 193–296.

[4] Burtt, H. E., 'Auditory Illusions of Movement,' J. of Exp. Psychol., 1917, 2, pp. 63–75.

[5] Benussi, V., 'Kinematohaptische Erscheinungen,' *Archiv für die gesamte Psychologie*, 1913, 29, pp. 385–388.

[6] Benussi, V., 'Kinematohaptische Scheinbewegung und Auffassungsumformung,' *Bericht VI. Kongres für experimentelle Psychologie*, Göttingen, 1914, pp. 31–35.

stimulus to the other in the direction of the actual temporal succession.

2. The longer the exposure of the stimuli the relatively shorter the interval which gives the optimal impression of movement.

3. The greater the distance between the stimuli the greater the optimal interval.

4. The greater the absolute intensity of the stimuli the shorter the optimal interval.

5. The greater the intensity the greater the optimal distance.

6. If the second stimulus is of greater intensity than the first the direction of the apparent movement is sometimes the reverse of the actual direction of temporal succession.

The above relations were investigated in the present case with tactual stimuli. The effect of varying degrees of difference in intensity between the two stimuli, which occupies a considerable portion of Korte's monograph, was not studied inasmuch as the reversal effect occurred much less frequently in the present case.

B. Apparatus and Method

An experiment of this sort necessitated some method of giving two similar tactual stimuli at different points of the skin, with control of the time relations, intensity of the stimuli and distance between them. It was also desirable to give as a check some stimulus approximating in its effect a continuous movement on the skin.

Ten brass rods with blunted points 2 cm. apart in a straight line were arranged to stimulate the back of the forearm. With the whole row actuated by solenoids in quick succession the effect was always much like a continuous movement, for any two adjacent points were within a "Weber's circle." This gave the check mentioned above. For most of the work only two solenoids were actuated on a given trial, the distance between stimuli depending on which solenoids were used.

The solenoids were mounted individually on narrow strips

of wood arranged to slide vertically side by side on a large wooden base. A piece of round magnet iron 2.5 cm. long formed the core and was suspended by a spiral spring from an adjusting screw above the solenoid. The lower end of this core contained a hole into which a brass rod of the same diameter as the core was dowelled. The lower end of this brass rod was pointed to give the tactual stimuli. The subject's arm was clamped palm downward in the rest of the Mosso's ergograph with the first and third fingers in the usual tubes to insure rigidity. The base to which were attached the strips of wood carrying the solenoids was mounted longitudinally above the arm. These strips could be moved up and down individually by set screws. The base also could be moved vertically so it was possible to adjust the 10 brass points to any desired distance from the skin. Any hairs in the vicinity of the brass points were of course removed. The iron part of the cores were usually set a few millimeters above center and the brass points adjusted to a distance of about 1 mm. from the skin. Thus the intensity of the stimulus did not vary appreciably with difference in length of time the current passed through the coils although it did vary with the intensity of the current. This latter was controlled by rheostats. The solenoids were actuated by either about 20 volts of storage batteries or by the direct 110-volt laboratory current with a rheostat in series.

A time-controlling mechanism was made from the 'Leipzig time sense apparatus' with a series of closely adjacent marginal contacts and a rather complicated switchboard. It was possible to actuate the 10 solenoids in rapid succession or any two in succession for desired lengths of time with any desired interval between. The space order (up or down the arm) could be reversed at will. The time mechanism was placed in an inner soundproof room. The experimenter and subject sat in an outer room at opposite sides of the table on which were the solenoids, switchboard and rheostats. The subject's arm rested comfortably in the apparatus which was so placed that the upright base was between the subject and the solenoids. Their action was thus invisible

to the subject but could be watched by the experimenter. The action of the solenoids was practically noiseless.

The subjects were not told of the nature of the experiment but merely that it was a study of tactual perception. At the beginning of each hour they were instructed: "When I say ready close your eyes and attend to the forearm and then tell what you felt."

The warning 'ready' was given approximately 1 second before the stimuli. On a given hour trials involving different relations of the given variables were presented in irregular orders. The main interest was in the effect of two successive stimuli, but trials were always intermixed in which a row of adjacent stimuli were given. It could thus be seen how the subject would confuse two discrete stimuli with the close approximation to continuous movement produced by a row.

The experiments were performed in the Harvard Psychological Laboratory in the academic year 1915–16. The subjects were graduate students or undergraduates of considerable psychological experience.

C. Results

I. *The Movement Illusion*

All four subjects who participated in the experiment yielded in many trials the illusion of movement. If the point of a rod was pressed on the skin of the forearm for a definite period and then after a brief interval another rod a few cm. distant pressed on the skin for the same length of time as the first, there was sometimes an impression of movement from the first point to the second, sometimes a single fused sensation and sometimes two discrete impressions. These effects varied with the combination of time, intensity and distance variables employed (*cf.* infra).

The (illusory) movement impression was variously characterized in the introspection of the subjects.

Subject *A* distinguished a *series* in which there appeared to be a number of stimuli in a row 'pretty close but possibly felt as separate,' from a continuous *flow* from one end to the other in which he 'could not tell the separate points.'

There was also at times an *arrow* at the end of the flow as if the movement went on in the air a short distance.

C noted three forms of the illusion: a *line* which was 'as if you pressed down a long object' either all at once or beginning at one end of the object; a *walk* in which there are a number of discrete points in succession; and a *roll* which is 'continuous movement.' He also occasionally reported a *loop*, the motion being in the air above the arm rather than actually upon the skin.

D described the apparent movement as a *snake effect.* He felt a number of intermediate points but also a connection. It suggested black points connected by a line; 'when they were close together the line was thick and when far apart it was thin.' The effect was best when all points appeared equidistant. A frequent report was three points in succession in the order *ABC* or *ACB.* In the latter case, with the middle point felt last, the snake effect appeared less readily.

F reported movement in which they 'skip along like playing four or five notes on a piano in succession.' It often suggested a little boy running along. He also noted frequently a sort of *loop* in the air between the ends. This 'blends into a sensation of movement.' This loop reported by *C* and *F* corresponds to Benussi's 'Bogenbewegung in der Luft.'[1] Benussi found that with a greater time interval the Bogen was greater and the movement worse. This seemed true for the present subject *F*.

II. *Exposure and Interval*

Korte found in his visual studies that the longer the exposure of the stimuli the relatively shorter the interval that yielded the optimal impression of movement. The writer found the same to be true with auditory stimuli.

To test this factor in the present case, intervals between stimuli of 15, 21, and 40 sigma were used with exposures of the same length or multiples thereof. The distance between the stimuli was constant at 12 cm.; the intensity was moderate, that given by a current of 1.5 amps. in the solenoids.

[1] *Op. cit.*, p. 32.

The results are summarized in Table I. The three successive columns represent results obtained with intervals of 15, 21 and 40 sigma respectively. At the left of each column are arranged vertically the exposures (in sigma) which were

TABLE I

EXPOSURE AND INTERVAL

(Distance 12 cm. Intensity 1.5 amps.)

Subject	Exposure	Interval 15	Exposure	Interval 21	Exposure	Interval 40
A			126	Two ends more discrete		
			105	Two ends	200	Two ends rather close together
			84	Series or two	160	Short series then other end
	45	Series or flow	63	*Series or flow*	120	Series three or four
	30	Two adjacent	42	Slight flow	80	*Four in flow or series*
	15		21	Single one in middle	40	Three in series
C			147	Two or three at each end		
	90	Ends successive	126	Five rolling with gap	240	*Four to six in succession*
	75	Ends successive	105	*Six or seven in succession (walk)*	200	Three or four in succession
	60	Four or five in succession (line) or two ends	84	One end then three rolling at other end	160	Three in succession
	45	*Four in succession (line)*	63	Three in succession	120	Ends successive but spread
	30	Ends successive	42	Ends successive or simultaneous	80	Ends successive but spread
	15	Ends successive or simultaneous	21	Ends successive or simultaneous	40	Ends successive
D			126	Group at ends		
			105	Two or three in series	200	Several at each end
			84	*Three or four with heavy line*	160	Three or four in series
			63	Three with thin line	120	Three or four in series
			42	Several simultaneous	80	*Four with snake effect*
					40	Three or four in series or ends
F	60	Ends	84	Ends		
	45	Ends	63	*Three or four*		
	30	*Three or four*	42	Two or three		
	15	Two adjacent	21	Two close together		

combined with the given intervals. In each section is stated briefly the phenomenon that generally occurred with that given combination of exposure and interval for the subject

indicated in the margin at the left. Each statement is based on from 5 to 10 trials and all results in the table were obtained on the same day. The check trials in which an actual series of touches was given by several successive solenoids are omitted. The results in this and subsequent tables are for only those trials in which two stimuli were given for equal lengths of time with an interval between. For instance we note that with the stimuli for 30 sigma each with an interval of 15 sigma between them, subject *A* generally reported two adjacent touches, and with stimuli for 42 sigma each and an interval of 21 sigma he noted usually a 'slight flow.'[1] Following across any horizontal row in the three columns we have exposure and interval in the same ratio. Thus in the example above 30 : 15 : : 42 : 21 : : 80 : 40. That report in each column which seems to indicate the greatest effect of the illusion is indicated in italics. Thus by noting the distribution of the italicized phrases for the different subjects any individual or general tendencies may be seen. The variables are arranged (in all tables) with the lowest values at the lower left corner so that an array of italics extending obliquely upward to the right indicates a positive relation between the variables and an array extending obliquely downward to the right indicates an inverse relation.

In the table it is evident that two of the subjects show a direct relation and two an inverse relation between the exposure and interval which give the best impression of movement. For instance with subject *A* an interval of 21 sigma is best combined with exposures of 63 sigma, 3 times as great, while a longer interval (40 sigma) is best combined with one only twice as great (80 sigma). That is, the longer the interval the relatively shorter the optimal exposure. *D* shows similar results. *C* and *F* manifest the opposite tendency but the latter's difference is very slight (a matter of 3 *vs.* 4 apparent stimuli) and perhaps negligible. On the whole the difference appears slightly in favor of the inverse relation. The longer the exposure of the stimuli, the relatively shorter the optimal interval. One may note further

[1] For explanation of individual notations, see p. 374–5.

that for a given interval the shorter exposures give an impression of simultaneity or fusion into a single touch, while increase of the exposure produces the movement illusion and still further increase an impression of discrete successiveness, a result in accord with Korte's visual findings.

III. *Distance and Interval*

Table II. summarizes the results with the distance between the points on the skin and the time interval between the stimuli constituting the variables. The exposures were constant at 60 sigma and the intensity at 1.5 amps. through the solenoids. The successive columns give results with the lengths of interval indicated. The rows give the different distances employed. The general tendency for each combination of interval and distance is indicated by a phrase as in Table I., and the report in each row showing the clearest evidence of the illusion is italicized, where such evidence is present. The results for *F* are indicated numerically rather than qualitatively. He usually gave the numbers of the points which he thought touched the skin (having adopted a scheme with No. 1 next the wrist, etc.). From three to six points were reported in each trial. These were simply added and averaged and the figure in the table represents the average number of successive stimuli reported per trial on the given combination when only two stimuli were actually given.

A glance at the italics shows in every instance a tendency toward a distribution obliquely upward to the right. That is, to obtain the optimal impression of movement the longer the interval between the stimuli the greater must be the distance. This result is the same as that found by Korte in vision. We may note further that for a given distance small intervals give a single point or simultaneity, while increase of interval gives movement and later discreteness and succession.

IV. *Absolute Intensity and Interval*

The intensity of the stimuli was varied from 1.8 amps. through the solenoids, which gave a moderate touch to 4.4

TABLE II

DISTANCE AND INTERVAL

(Exposure 60 sigma. Intensity 1.5 amps.)

Subject	Distance	Interval					
		15 Sigma	30 Sigma	45 Sigma	60 Sigma	75 Sigma	105 Sigma
A	16 cm.		Ends or flow at each end	Ends	Ends	Ends	Ends
	12 cm.		1 spread out	2 adjacent with arrow	2 or 3 adjacent	*3 in series*	
	8 cm.		1 or short flow	1 or short flow	*Flow*	Flow or series	Series of 3 or 4
A (later)	16 cm.	Ends	Ends	Ends	Ends	Ends	
	12 cm.	Short series or flow	Series or 2 adjacent	*Flow or series*	Two	Two	
	8 cm.	*Flow*	Series	Series or 2 adjacent	2 adjacent	2 adjacent	
C	16 cm.	Ends successive or simultaneous	Ends or series	Ends or series simultaneous	Ends each spread	Ends each spread	Ends
	12 cm.	Ends	Ends	2 or 3 successive or simultaneous	*4 in succession*	*4 in succession*	Ends spread
	8 cm.	Several in succession or simultaneous	3 in succession	*3 or 4 in succession*	*3 or 4 in succession*	3 in succession	3 in succession
D	16 cm.	Ends	Ends	Ends	*Slightly snaky*	Ends	
	12 cm.	Several simultaneous	Several simultaneous	*Snake*	*Snake*	2 or 3 in series	
	8 cm.	One	One	One	One	2 near together	
F	16 cm.	3.2		*4.8*		4.0	
	12 cm.	*7.2*		5.8		4.8	
	8 cm.	*5.0*		4.2		4.4	

(Exposure 120 sigma. Intensity 1.5 amps.)

		40 Sigma	80 Sigma	120 Sigma	160 Sigma
C	16 cm.	Discrete groups	*Discrete groups walking or rolling*	Ends slightly spread	Ends slightly spread
	12 cm.	*5 or 6 rolling*	4 or 5 walking or line	Discrete groups	Discrete groups walking
	8 cm.	Several walk	Several walk or roll	Several walk with gap	Several walk

amps., a stimulus so intense as to be startling. In the separate columns of Table III. are the results with different intensities. The rows give the intervals. The exposure was held constant at 60 sigma and the distance at 12 cm. The

TABLE III

INTENSITY AND INTERVAL

(Distance 12 cm. Exposures 60 sigma)

Subject	Interval	1.8 Amps.	2.1 Amps.	2.4 Amps.	3.2 Amps.	4.4 Amps.
A	75 sigma	Two adjacent	Two adjacent	Ends	Ends	Ends
	45 sigma	*Two adjacent or flow or arrow*	*Two adjacent or flow or series*	Two or series	Two quite close	Two farther apart
	15 sigma		Series or flow	*Flow*	Two adjacent	Two farther apart
C	75 sigma		Several in row	*Several in row or line*	Ends or row with loop	Ends or row with loop
	45 sigma		One or two	*Several in row sometimes loop*	Several in row	Several in row or ends
	15 sigma		Several in row or one	Several in row	*Series with loop or lot of motion*	Series or ends
D	75 sigma		2.0	2.2	3.4	*3.8*
	45 sigma		1.6	2.4	3.0	*3.4*
	15 sigma		1.6	2.0	2.6	*3.8*
F	75 sigma		4.2	*4.8*	3.8	4.4
	45 sigma		4.4	4.6	*5.2*	*5.2*
	15 sigma		4.4	6.6	5.6	*7.0*

general tendency shown on each combination is briefly indicated as in previous tables and the phrase in each row which indicates the optimal illusion italicized. The results for *D* and *F* are recorded numerically as in Table II.—the average number of touches felt per trial.

The italics are distributed with three of the four subjects in a direction obliquely downward toward the right. With *D* the greatest movement occurred always with the greatest intensity. Thus the general tendency of the majority is for a greater intensity to require a lesser interval to produce the movement impression. Further, for the two subjects with qualitative results, with a given interval as the in-

tensity increases the impression passes from that of a single touch, through a row or motion to two discrete touches. Both these facts agree with those found by Korte.

V. *Distance and Intensity*

In Table IV. the distance between the stimuli and the intensity of the stimuli were the variables. The times of exposure and interval were constant, in most cases 60 and 30 sigma respectively. In a few series other times were used, as indicated. The successive columns indicate intensity of the stimuli in terms of the current in the solenoids. The rows for any subject indicate the distance between the stimuli. The most frequent report for each combination of distance and intensity is indicated as previously. In a given row the report which shows the most marked movement illusion is italicized. *F*'s results are given in the form of the previous tables, *i. e.*, the average number of touches felt per trial.

The array of italics for the various subjects shows a rather general tendency toward a distribution extending obliquely upward toward the right, *i. e.*, with a greater distance between stimuli a greater intensity affords a better illusion of movement. This too is in accord with the results of visual experiments. Further, for a given distance, as the intensity varies from weakest to strongest there is somewhat of a tendency for the effect to pass from unity through movement to discreteness, and for a given intensity increase of distance tends to produce increase of discreteness, a result similar to Korte's. In this latter connection we may note Benussi's finding that with distances greater than 16 cm. on the forearm there was an individualizing of the ends following the superficial movement.

VI. *Unequal Intensity*

In all the experiments up to this point the two successive stimuli on a given trial were equal in intensity. This factor was now varied as had been done in the visual and auditory experiments. With exposures of 60 sigma and interval of

TABLE IV

DISTANCE AND INTENSITY

(Exposures 60 sigma. Interval 30 sigma)

Subject	Distance	1.2 Amps.	1.4 Amps.	1.5 Amps.		
A	12 cm.	Ends	*Flow or series*	*Flow or series*		
	8 cm.	*Flow or series*	Series	Ends or series		
C	12 cm.	One or both ends	*One end rolling*	Ends sometimes spread		
	8 cm.	*Three rolling*	Three rolling or walk	Ends sometimes spread		
D	12 cm.	Ends	*Ends or several simultaneous*	Ends		
	8 cm.	One or snake	*Good snake*	Two or snake		
F	12 cm.	4.4	5.0	3.8		
	8 cm.	4.2	4.4	4.4		
		2.1 Amps.	2.3 Amps.	2.8 Amps.	3.6 Amps.	4.4 Amps.
A	16 cm.	Ends close		Ends		Ends
	12 cm.	Ends or flow or series		*Flow or series*		Ends
	8 cm.	*Ends or flow*		Ends		Ends
D	16 cm.	Ends or one		Ends		*Ends or series*
	12 cm.	One or series		*Series*		*Series*
	8 cm.	One or two		*Long series*		Series
C	16 cm.	Ends	Ends	Ends	*Ends with connection*	
	12 cm.	Two adjacent	*Two close, with loop*	Two at one end	Two adjacent	
	8 cm.	One or two	*3 or 4 successive*	One or two	One or two	

(Exposures 60 sigma. Interval 45 sigma)

		2.1 Amps.	2.3 Amps.	2.8 Amps.	3.6 Amps.
C	16 cm.	Ends or one	Ends with line on skin	Ends with or without loop	*Ends with loop*
	12 cm.	One or two	One or two	*2 or 3 with loop*	Two connected or with loop
	8 cm.	One	*One or two with loop*	One or two	One

30 sigma, with stimuli 8 or 12 cm. apart and with the intensity of the first stimulus 1.5 amps., that of the second was increased from 0.5 to 2 amps. It was varied irregularly through various intensities and trials were intermixed in which both were equal.

Table V. gives the per cent. of trials in which the apparent movement occurred in the direction the reverse from that of the actual temporal succession. The first column gives the per cent. of trials in which the reversal effect occurred when the second stimulus was more intense (including all variations

TABLE V

UNEQUAL INTENSITIES

(Exposures 60 sigma. Interval 30 sigma. Distance 8 or 12 cm. Intensity 1.5 amps.)

Subject	Per Cent. of Trials Yielding Reverse Movement: Second Stronger	Equal Intensity
A	10.5	0
C	21.4	0
D	32.0	0
F	13.1	0

of intensity); the second column gives the same figures for the trials with equal intensity of the two stimuli. The results were obtained on several days and are based on some 50 trials of each sort for each subject.

It is evident that whereas with equal intensity of the stimuli the reversal effect never occurs, with the second stimulus more intense by various amounts the effect occurs in from 10 to 32 per cent. of the trials with the different subjects, an average of 19 per cent. This per cent. is much less than that found by the writer in the auditory illusions.

No attempt was made to correlate the results with the extent of the intensity difference or to study other aspects of the reversal effect (Korte's delta movement) as was done in the visual studies for the simple reason that the reversal effect occurred so much less frequently. However it is significant that the effect does exist there just as it does in vision and audition.

D. CONCLUSIONS

The above experiments investigated in the tactual illusion of movement the most salient of the factors noted by Korte in the similar visual illusion. The following are the principal results:

1. Two punctate tactual stimuli on the forearm for equal lengths of time separated by a discrete time interval and a

few centimeters apart, yield under certain conditions of time, distance and intensity an impression of movement from one point to the other in the direction of the actual temporal succession.

2. To produce the optimal impression of movement the following relations are required between the different variables under consideration:

(*a*) The relation between the length of exposure of the two stimuli and the time interval between them is not especially clear cut but on the whole is inverse—the longer the exposure the relatively shorter the interval.

(*b*) The greater the distance between the stimuli the greater the time interval.

(*c*) The greater the intensity of the stimuli the less the interval.

(*d*) The greater the distance between the stimuli the greater the intensity.

3. If the intensity of the second stimulus is greater than that of the first the illusory movement is sometimes produced in the reverse direction.

The above results are quite similar to those found by Korte in the visual illusion of movement. They are also the same as those found by the writer in a preliminary study of auditory illusions as far as that investigation was carried.

It seems a significant fact that the same illusion of movement is found in vision, audition and touch and that the same laws as to the relation between the variables involved hold to a considerable degree in the three fields. This points of course to a central origin. Whether it is due as Wertheimer claims to a physiological 'Kurzschluss' between cortical areas corresponding to the two stimuli at different points in space is an open question. Such a correlation between points of external space and points of the brain is perhaps plausible in vision and touch. The writer has pointed out[1] that such a correlation is highly improbable in audition and that hence a different explanation is necessary, and inasmuch as the three sense departments show the same phenomena it is better to adopt a single theory applicable to all three.

[1] *Op. cit.*, p. 75.

The writer has suggested (*op. cit.*) that the illusion is due to a continuity of motor impulse, and such a theory would be applicable to touch as well as to vision and audition. A tactual stimulus on the forearm produces an impulse to make some motor adjustment to bring that region into the focus of attention, presumably an impulse to turn the head or eyes toward that point. This would have, of course, a biological significance. A second stimulus, farther up the arm for example, gives a second impulse, presumably in the same centers of the motor cortex, to turn the head or eyes still farther. If the interval is of such length that the second impulse supervenes just when the first is about exhausted this continuity of motor impulses gives the movement illusion. When the second follows too quickly the two are not differentiated sufficiently and there is fusion and if the second is too late there is discontinuity and discreteness. Benussi found it impossible to obtain certain of the phenomena when the stimuli were applied to the forehead. This fact is suggestive inasmuch as the usual motor adjustments of the head to bring that region into the focus of attention are impossible.

The various relations between the different variables necessary to produce the optimal illusion are as explicable by such a theory as by the Kurzschluss theory. If we assume that intensity, distance and length of exposure are related positively to the temporal facilitation of the motor impulses and to their consequent sooner subsidence, all the above phenomena can be readily explained. The writer reiterates his suggestion that the action theory can account for the movement illusion which occurs similarly in vision, audition and touch.

ASSOCIATION–REACTION AS A TEST OF LEARNING[1]

BY KNIGHT DUNLAP

Since 'mental' learning involves primarily the formation of new associations, it is obvious that there is the possibility of testing the learning by some form of the *association reaction*. Moreover, since we can control the associative recall by the proper instructions, it would seem possible to determine the amount of intelligent learning (intellectual association) which has occurred, excluding rote-learning, or association by contiguity: and, by careful selection of controls, it might be possible to determine whether the student has acquired the proper habits of thinking in a given subject-matter, as discriminated from a mere acquaintance with details. Although theoretically possible, the practicability of such a learning-test has remained to be decided by experimental work. I have had such experimentation in mind for some time, and have this year, through the assistance of eight members of my 'teachers' course' in introductory psychology, been enabled to make a beginning in the work. Although a mere beginning, the results have been so decidedly encouraging, that it seems advisable to describe the tests for the benefit of any who may be interested in the same line of work.

The value of a test of the sort undertaken depends on the student's attitude towards it, especially if it is desired to compare the results with the results of ordinary written examinations and class-work. Hence, I announced to my class that any members who were willing to let their standings in the second half-year's work depend on the results of an association-test, would be permitted to take this test in lieu of the regular written examination. Nine students volunteered under this condition, which makes the results com-

[1] From the Psychological Laboratory of the Johns Hopkins University.

parable with ordinary ratings of these students, in so far as the seriousness of their attitudes was concerned.

With careful reference to the assigned text, and the lectures, of the semester, the following list of stimulus words was prepared, each word being accompanied by appropriate control instructions. The significance of these words and instructions depends, of course, on the way in which the materials were presented in class, and hence cannot be adequately explained here. The instructions for each stimulus word were read to the reactor just preceding the giving of the word; and because the instructions were not adequate, they were supplemented by additional extemporaneous instructions, sufficient to make clear to the reactor the relations involved in the control; yet without giving him more information than is actually intended by the written instructions. Extemporizing in this way is of course a fault, and inevitably leads to unfairness, however carefully it may be done. With more carefully selected words and more skillfully prepared instructions, supplementation and its evils can be excluded.

It was made clear to the reactors that the *response* under any given control need not satisfy the grammatical relations of the instruction so long as it clearly indicated the matter which, under the instructions, was demanded by the stimulus word. It was explained that in every case a satisfactory response could be given in one word, but, that in case a logically valid response consisting of a phrase were given, it would be accepted. The reactor was instructed that his primary duty was to give the correct response, but that it was to be given as quickly as possible; that if he thought of the correct answer after giving an incorrect one, he should give the correct one also.

The reaction times were taken by means of the Johns Hopkins chronoscope[1] and the voice keys[2] of the new (smaller) model. The chronoscope, running on a 50 vibration fork, read in units of 2σ.

[1] J. of Experimental Psychology, Vol. 2, No. 3.

[2] Psychol. Rev., 1913, **20**, 250–253.

INSTRUCTIONS AND STIMULUS WORDS

1. Reply with the species under the genus I name: that is, give one sort of whatever I name. (Examples: Fruit—Peach; Crime—Murder.) *Consciousness.*

2. Reply with another member of the same class to which belongs whatever I name: give something of the same sort. (Examples: English—Italian; Chair—Table.) *Thyroid.*

3. Same instructions. *Ideo-motor.*

4. Reply with the name of the gland which controls the process I name. *Growth.*

5. Reply with the essential factor involved in that which I name: that which makes it what it is. *Memory.*

6. Same instructions. *Will.*

7. Same instructions. *Judgment.*

8. Reply with the emotion which, in regard to the inclusion of certain elements, is the opposite of the one I name. *Religious.*

9. Reply with the affection (affective content) which in its occurrence or behavior is characterized by that which I name. *Regression.*

10. Reply with that which in regard to *time* is the opposite of what I name. (Example: Yesterday—Tomorrow.) *Anticipation.*

11. Reply with that which is essential to that which I name and whose existence is demonstrated by the latter. *Recall.*

12. Reply with the most important factor involved in that which I name. *Emotion.*

13. Reply with one of the stages in that which I name. *Learning.*

14. Reply with one of the two sorts of that which I name. *Ego.*

15. That which I name depends on a peculiarity of consciousness which is exhibited in all perception. Reply with this peculiarity, or the more general manifestation of this peculiarity. *Rhythm.*

16. Reply with the opposite of that which I name (not the verbal opposite, but that which is actually opposite). *Instinct.*

17. The most fundamental psychological problem is in the relation between that which I name and a second fact. Reply with this second fact. *Reflex.*

After comparing the results of the eight reactors, it seemed possible to evaluate them according to the following rules.

1. Each correct response, given in 900 to 1,000 units, was assigned ten points credit.

2. For a correct reply in 800 to 900 units, one point; in 700 to 800, two points; 600 to 700, three points; and 500 to 600, four points additional credits were given.

3. For a delay from 1,000 to 1,200 units, one point; 1,200 to 1,500, two points; 1,500 to 2,000, three points; 2,000 to 3,000, four points; 3,000 to 5,000, five points, were subtracted.

4. Answers not completely satisfactory, but having a real bearing on the requirements were evaluated at less than ten. (In a properly prepared test, no such cases should arise.)

5. The total credits so obtained were divided by the 'normal' of 170.

The responses obtained from the nine subjects are given below, together with the credits assigned for the normal time (900–1,000 units). Where no credit is noted, it was zero. The numbers in parentheses indicate the number of reactors giving the same response: where no such number is given, 1 is understood.

RESPONSES

1. 'thought' ten points; 'present' ten points; 'content' (2); 'sensation;' 'reflex-arc;' 'sub-consciousness;' *no response* (2).

2. 'adrenal' (3) ten points; 'pituitary' (4) ten points; 'thymus' ten points; 'gland.'

3. 'perceptual' (2) ten points; 'volitional' ten points; 'response;' 'ideational;' 'motor;' 'gland;' *no response* (2).

4. 'thyroid' (6) ten points; 'pituitary' (3) ten points.

5. 'recognition' (3) ten points; 'retention' (2); 'perception' (2); 'reproductive imagination;' 'recall.'

6. 'desire' (4) ten points; 'volition—desire' eight points; 'action' six points; 'volitional' (2); 'consciousness.'

7. 'concept' (5) ten points; 'consciousness' (2); 'thought;' *no response.*

8. 'atheist' 'doubtful;' 'non-religious;' 'anti-religious;' 'sacrilegious;' 'secular;' 'faith;' *no response* (2).

9. 'conation' (2) ten points; 'conative' ten points; 'repugnance' ten points; 'progression;' 'retrogression' 'association;' *no response* (2).

10. 'memory' ten points; 'realization;' 'disappointment;' 'hope;' 'future;' 'looking backwards;' *no response* (3).

11. 'retention' (2) ten points; 'association' (2) eight points; 'memory;' 'perception' (2); 'perceive;' 'experience.'

12. 'cognitive elements' ten points; 'bodily sensation' (3) ten points; 'object' (meaning 'the cognitive element') ten points; 'feelings—physiological activities' ten points; 'feeling' ten points; 'hedonic tone' ten points; 'pleasure.'

13. 'imitation' (8) ten points; 'recall—memory.'

14. 'self' (3) ten points; 'conscious;' 'sub-conscious;' 'selfish—egotistical;' *no response* (3).

15. 'specious present' (5) ten points; 'periodicity—specious present' ten points; 'time;' 'feeling;' *no response.*

16. 'habit' (3) ten points; 'learning' (2) ten points; 'consciousness;' 'volition;' 'memory;' 'emotion.'

17. 'consciousness' (5) ten points; 'integration' ten points; 'instinct;' 'action;' *no response.*

As may be inferred from the list, the test was somewhat too difficult. The reactors graded from 90 per cent. to 30 per cent., when they should have graded, according to classwork and previous tests, from 100 per cent. to 60 per cent. The 'order of merit' and relative standings in the test were however what might have been expected from previous work of the students, with one exception: one student who made the low mark (30 per cent.) confessed afterwards to have slighted the reading in the second semester, and to have done no reviewing; this was done with full intention, from curiosity to 'see what the test would show.' Two of the students had reviewed with especial regard to the test; attempting to strengthen word-association in the various topics which they conjectured might be of service, basing their conjecture merely on a general acquaintance with the association-reaction method. These conjectures of course gave the students no help at all on the actual test.

The results of the test brought out sharply the difference between topics which had been thoroughly discussed in class, and those (*e. g.*, 9) which had been slighted, and also difference in difficulty in the topics themselves. As regards the individual differences of the reactors, the test alone does not show whether these were due to differences in amount of work done, or to difference in the ability of the student to grasp the vital points, with an average amount of effort. It is probable, however, that a more carefully prepared test would differentiate these causes. One thing which stands out sharply is that 'bluffing' is entirely excluded: the student makes a good or poor showing, according as he 'knows' or does not know the subject, whatever the contributing cause of his grasp or lack of grasp.

As in all association-tests, the most vital point is in the construction of the word-tests, and the instructions. A special difficulty, which must be solved experimentally, is the derivation of a just and satisfactory method of evaluating the results of the test.